2017

A FORENSICALLY-ENABLED IAAS CLOUD COMPUTING ARCHITECTURE

Alqahtany, Saad

A FORENSICALLY-ENABLED IAAS CLOUD COMPUTING ARCHITECTURE

By

SAAD SAID ALQAHTANY

A thesis submitted to the University of Plymouth

in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Electronics and Mathematics

Faculty of Science & Engineering

January 2017

Abstract

A Forensically-Enabled IaaS Cloud Computing Architecture

Saad Said Alqahtany

Cloud computing has been advancing at an intense pace. It has become one of the most important research topics in computer science and information systems. Cloud computing offers enterprise-scale platforms in a short time frame with little effort. Thus, it delivers significant economic benefits to both commercial and public entities. Despite this, the security and subsequent incident management requirements are major obstacles to adopting the cloud. Current cloud architectures do not support digital forensic investigators, nor comply with today's digital forensics procedures – largely due to the fundamental dynamic nature of the cloud.

When an incident has occurred, an organization-based investigation will seek to provide potential digital evidence while minimising the cost of the investigation. Data acquisition is the first and most important process within digital forensics – to ensure data integrity and admissibility. However, access to data and the control of resources in the cloud is still very much provider-dependent and complicated by the very nature of the multi-tenanted operating environment. Thus, investigators have no option but to rely on the Cloud Service Providers (CSPs) to acquire evidence for them.

Due to the cost and time involved in acquiring the forensic image, some cloud providers will not provide evidence beyond 1TB despite a court order served on them. Assuming they would be willing or are required to by law, the evidence collected is still questionable as there is no way to verify the validity of evidence and whether evidence has already been lost. Therefore, dependence on the CSPs is considered one of the most significant challenges when investigators need to acquire evidence in a timely yet forensically sound manner from cloud systems.

This thesis proposes a novel architecture to support a forensic acquisition and analysis of IaaS cloud-base systems. The approach, known as Cloud Forensic Acquisition and Analysis System (Cloud FAAS), is based on a cluster analysis of non-volatile memory that achieves forensically reliable images at the same level of integrity as the normal "gold standard" computer forensic acquisition procedures with the additional capability to reconstruct the image at any point in

time. Cloud FAAS fundamentally, shifts access of the data back to the data owner rather than relying on a third party. In this manner, organisations are free to undertaken investigations at will requiring no intervention or cooperation from the cloud provider.

The novel architecture is validated through a proof-of-concept prototype. A series of experiments are undertaken to illustrate and model how Cloud FAAS is capable of providing a richer and more complete set of admissible evidence than what current CSPs are able to provide. Using Cloud FAAS, investigators have the ability to obtain a forensic image of the system after, just prior to or hours before the incident. Therefore, this approach can not only create images that are forensically sound but also provide access to deleted and more importantly overwritten files – which current computer forensic practices are unable to achieve. This results in an increased level of visibility for the forensic investigator and removes any limitations that data carving and fragmentation may introduce.

In addition, an analysis of the economic overhead of operating Cloud FAAS is performed. This shows the level of disk change that occurs is well with acceptable limits and is relatively small in comparison to the total volume of memory available. The results show Cloud FAAS has both a technical and economic basis for solving investigations involving cloud computing.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This thesis would not have come to successful completion without guidance and support from my Director of Studies Professor Nathan Clarke who has been very supportive of me. I would like to express my special appreciation and thanks to him. I really owe much of my success to his tireless effort through the Ph.D. process.

Thanks must also go to my other supervisors, Professor Steven Furnell and Professor Christoph Reich who have spent a lot of time proofreading papers and my thesis, in addition to providing helpful experience and guidance throughout my studies.

I'm very much indebted to my mom and dad who call me and support in every single week during my Ph.D. research journey.

I'm very much indebted to my wife, Norah – who supported me in every day during my Ph.D. research journey and spent sleepless nights with my children, although she is over busy with her Ph.D. research.

My thanks also go my best friend – Abdulwahid for his support and for the motivating ideas and thoughts he provided during my Ph.D. journey. I would also like to thank my colleagues who participated in my research namely: Ayad Al-Adhami, Saud Alotaibi, Abdulrahman Alrubian and Leith Abed. I also owe a great debt of gratitude to all the experts who have contributed in the evaluation of this research.

I dedicated this thesis to my children Abdullah, Naif, and Reema.

Authors Declaration

At no time during the registration for the degree for Doctor of Philosophy has the other been registered for any other University award.

This study was financed with the aid of a scholarship from the Kingdom of Saudi Arabia.

Relevant seminars and conferences were attended at which work was often presented and several papers prepared for publication.

Word count of the main body of thesis: 44,050 words

Signed ……Saad……...

Date ……04/06/2017……...

1 Introduction and Overview

While security has frequently been an “after-thought” in Cloud Computing, digital forensics has been an “after-after-thought” (Ruan & Carthy, 2013). There is a high level of demand on the forensic-aware tools for the Cloud Service Provider (CSP) and the clients to conduct a forensic investigation in the cloud environment (Zawoad & Hasan, 2013a). This research presents a novel approach to forensic acquisition and analysis within an Infrastructure as a Service (IaaS) cloud model that both meets existing standards and shifts the control of the forensic analysis to the cloud user (who actually owns the data) rather than the cloud provider.

1.1 Introduction

Cloud computing technologies have significantly changed the way in which organisations implement the building of their information technology infrastructure (CSA, 2016). Cloud computing offers significant economic benefits to users by providing a highly scalable infrastructure and pays as you go services at low cost and on-demand computing (Banas, 2015). Using the notion of ‘pay as you use’ provides an economic solution for organisations and makes a clear indication that the cloud has quickly become an essential ingredient of modern IT (Florentine, 2016). Most of the traditional IT hardware including servers, networks, routers, and switches are not hosted and managed by the user’s organisation; organisations are outsourcing them to third party providers. The cloud provider replaces them with virtualised, remote on-demand software services (Grispos et al., 2011). The payment model is a pay per use method which allows organisations to focus on their business with minimal effort placed upon building, managing and maintaining their IT requirements.

Khajeh-hosseini & Greenwood (2010) found that organisations which moved their IT infrastructure from an outsourced data centre to a cloud provider such as Amazon could obtain

a 37% cost saving and would eliminate 21% of the support calls for their system. The use of cloud computing is growing and Gartner suggests this growth will increase to become the bulk of new IT spend by 2016 (Gartner, 2013).

Moreover, according to the latest research conducted by Cloud Industry Forum (CIF) the United Kingdom states has more than 70 % of the business use at least one cloud service and 80% of current users will increase their spend (Cloud Industry Forum, 2012).

Cloud computing will deliver computing resources as utilities in the same way that familiar utilities are delivered, and just like gas, water and electricity, the user will access services on demand and pay for their use. Due to these promising characteristics and financial benefits, cloud computing has become increasingly attractive for both commercial and public entities. Gartner pinpointed that Cloud services are most disruptive forces of IT spending ever of the digital age, by 2020 "Cloud Shift" will affect more than $1 Trillion in IT spending (Gartner, 2016).

Although significant opportunities are clearly offered by cloud computing for organisations, security issues are ranked as the single greatest challenge of cloud computing (Kuyoro et al, 2011). According to a survey conducted by Dell (2014), the issue of security is listed as the top concern of cloud adoption. Issues surrounding the security and subsequent incident management requirements in the cloud are frequently ignored leading to devastating consequences (Josshua, 2012).

With stolen credentials, attackers can often access critical areas such as cloud computing services, allowing them to compromise the confidentiality, integrity and availability of those services. In 2014, Code space’s Amazon AWS account was compromised. The hacker removed most of the company's data, backups, machine configurations and offsite backups were either

partially or completely deleted. This cloud-based attack left Code Spaces unable to operate and it went out of the business (CSA, 2016).

Furthermore, Anthem Inc., the second-biggest U.S. health insurer, was the target of a very sophisticated external cyber attack in 2015; more than 80 million customer records were stolen from the corporate network. Such information includes names, birthdays, medical IDs, Social Security numbers, street addresses, e-mail addresses and employment information, including income data (Weise, 2015). The investigative team revealed that the hackers downloaded the data while a third-party cloud service transferred the huge data store from the company's network to the public cloud (CSA, 2016).

Unfortunately, enterprises are in a rush to adopt cloud computing and move ahead without the implementation of security and forensics capabilities at the initial stage of the cloud's architectural design (Srinivasa et al., 2011). While security has frequently been an "after-thought" in new technology such as the cloud, digital forensics has historically been an "after-after-thought" (Ruan & Carthy, 2013). When the evidence resides in the cloud, new challenges exist on how to apply current digital forensics procedures. These challenges are novel and unique to the cloud and not encountered in traditional digital systems. This is due to the unique combination of characteristics that cloud computing introduce, including; on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service (Mell & Grance, 2011).

The security breach situation appears to be getting worse as cyber hackers have already taken advantage of cloud features. The term 'crime–as–a-service' has been introduced and there are many cybercrimes that are available for sale (Daly, 2012). Hackers can offer illegal services at competitive prices such as launching huge denial of service attacks. For example, the owner

of a Russian payments firm (Chronopay) allegedly hired two brothers to attack a competing payment processor firm (Assist) (Krebs on Security, 2013).

Cloud-computing services are as easy and convenient for customers as they are a target for an intruder. Intruders know that a single breach can lead to a treasure trove of material and they are skillful to breach various cloud services. For example, the breach of Apple's iCloud service, copying risqué photos of many celebrities and leaking them online (Steinberg, 2014). In another example is a massive cloud computing breach which occurred in 2011 affected a number of the largest entertainment companies and cloud service providers. It was the second-largest online data breach in the history of U.S. and it cost the intruders three pennies an hour to rent the Amazon servers that launch the highly sophisticated incursion (Galante et al, 2011).

Conducting a digital investigation without forensics by design is an expensive process. Fannie Mae (The Federal National Mortgage Association) spent approximately 9% of its total annual budget on the discovery of electronically stored information for litigation purposes and still failed to meet discovery deadlines. According to a survey on cloud and eDiscovery disseminated to organisations that are using cloud-based solutions, 26% responded that they do not have an eDiscovery plan in place and 58% responded that they do not even know if a plan exists (Murphy, 2011). This means that in the case of litigation and investigation, such organisations will be left scrambling in a reactive manner to collect information from the cloud; leading to greater cost (Ruan, 2013).

Birman (2010) stated that "we are still in the early days, year 1 for cloud computing may be closer to the year 2015". As the cloud is in an infancy stage of development, there are therefore many promising opportunities to overcome all the difficulties faced by digital forensics investigators in the cloud environment. Moreover, to date, there is still a lack of studies focussing upon the current state of cloud forensics tools, techniques, and processes in order to

obtain evidence from the cloud environment in a forensically sound manner. Furthermore, security and forensics practitioners agree that frameworks of processes that need to stand up in a court of law and provide admissible evidence have yet to be developed (Zimmerman & Glavach, 2011).

1.2 Aim and Objectives

The aim of this research is to develop a novel approach to forensic acquisition and analysis within an Infrastructure as a Service (IaaS) cloud model that both meets existing standards and shifts the control of the forensic analysis to the cloud user (who actually owns the data) rather than the cloud provider. This aids in reducing the complexity of undertaking the data acquisition process, the requirement for the CSP's active involvement and any modification to the underlying architecture of the CSP.

To achieve the aforementioned aim, the following objectives have been set:

- Establish a current state-of-the-art understanding of cloud computing architectures and digital forensics, with a view to identifying and assessing the necessary attributes required to enable forensics procedures in a cloud environment.
- Explore the new challenges that cloud computing brings to digital investigation processes and seek to evaluate the extent to which current forensics techniques can be applied to cloud computing.
- Develop a cloud computing test bed from which subsequent analysis and evaluation activities can be undertaken.
- Establish the current state-of-the-art key issues of cloud forensics, including the critical criteria for cloud forensic capabilities that will help to establish the requirements for a novel model that ensures a full forensics capability within cloud environments.

- Develop, implement and evaluate the proposed model to ensure it is an appropriate trade-off between managing the security requirements of consumers and the associated costs.
- Evaluate the feasibility of the designed approach by seeking opinions and feedback from experts in the field including practitioners and researchers.

1.3 Thesis Structure

The remainder of the thesis is structured as follows: Chapter 2 provides a background on cloud computing, including its definition, the most common services and deployment models and its essential characteristics. Cloud computing consists of different deployment and service models, which are then elaborated. Furthermore, the cloud architecture is presented along with technologies that are related to cloud computing.

Chapter 3 presents an overview of digital forensic science, seeking to summarise current digital forensics capabilities, tools and techniques. It also looks at current state-of-the-art knowledge, with respect to digital forensics, in order to identify the appropriateness of the standards and tools necessary to undertake digital investigations in a cloud environment. The main methodologies used when conducting digital investigations are discussed alongside the critical digital forensic issues that have not been addressed thoroughly in the literature.

Chapter 4 introduces the forensics artifacts in cloud environments, current cloud forensic frameworks, followed by its challenges and available technical solutions. It also matches identified solutions with the addressed challenges as well as identifying the open problems in the domain based on the detailed literature review. These open problems were chosen as the scheme orientation for the offered research.

Chapter 5 introduces a novel approach that addresses the main research gap in cloud forensic fields in order to support forensics acquisition and analysis within an IaaS service model. It starts with the basic development, which focuses on the concept of implementing an agent-based approach that sits on each of the customer's VMs. Such an approach needs to be experimentally proven in order to illustrate the functionality of the proposed approach. Thus, three experiments are conducted in Chapter 5 with the aim of investigating the approach's admissibility, efficiency, and feasibility. This chapter provides a better understanding of both the technical implications resulting from such a system, regarding the day-to-day operation of a cloud system and the financial costs.

The broad architecture of the Cloud FAAS is presented in Chapter 6, with an accompanying description of each of the key components. In order to provide an operational system, the functionality of various components is discussed, as well as the contribution of each to the complete system. Besides this, the technical evaluation and cost–benefit analysis of the model are discussed.

In Chapter 7, an expert-based evaluation is presented. Experts from different areas, including academics and industry practitioners, are interviewed with the aim of exploring a further area of development of the proposed system, along with identifying the system's strengths and weaknesses from different perspectives.

Finally, Chapter 8 summarises the conclusions arising from the research, highlighting the key achievements and limitations. This chapter also contains a discussion on areas for future research. The thesis also provides a number of appendices in support of the main discussion, including code listings and a number of published papers arising from the research.

2 Cloud Computing

Cloud computing has become a new paradigm in information technology. Cloud computing is likely to become an evolutionary point in the computing era due to the promising technological and economic opportunities it provides for public, private and government organisations (Daryabar et al, 2013). People do not have to think about infrastructure investment, management and maintenance issues. They just need to have Internet access anywhere and at any time to get cloud services. This chapter introduces cloud computing, its service models and deployment models. Furthermore, it presents the cloud architecture along with technologies that are related to cloud computing.

2.1 Definition of Cloud Computing

Google's CEO Eric Schmidt used the term 'Cloud' to describe the business model that provides services via the Internet in 2006. Since then, the term 'cloud' has started to gain popularity and has also been used as a marketing term to present many different ideas (Zhang et al., 2010). Yet, there is no single standard definition of cloud computing and this lack of standardisation can lead to confusion of what cloud computing is, although some research was carried out to solely confirm a universal standard definition of cloud computing. For example, Vaquero et al. (2009) have compared more than 20 definitions in the aim of extracting a consensus definition. However, Jim Reavis, executive director, Cloud Security Alliance stated that "NIST has provided valuable leadership in defining cloud computing and in characterising its various opportunities and challenges" (Brown, 2012). Furthermore, the Cloud Security Alliance (2009) aligned with National Institute of Standard Technology's definition because of it's general acceptability. NIST (2011) defined cloud computing as:

"… a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"

It is essential to identify cloud computing characteristics in order to understand their impact on security and digital forensics. There are five essential characteristics of cloud computing identified by NIST (Mell & Grance, 2011):

1- Demand self-service: Cloud computing capabilities such as service time and storage capacity can be provisioned to the cloud consumer as needed unilaterally by the consumer without human interaction from the cloud provider service.
2- Broad network access: Services hosted in the cloud are web-based. Thus, it is easy to access them through standard mechanisms and the variety of devices with Internet connectivity including PCs, laptops, cell phones and PDAs.
3- Resource pooling: the provider's resources are pooled and a multi-tenant model is used in order to dynamically assign the resources to multiple customers. However, the consumer can specify resource's location at a higher level of abstraction such as country or state but they do not exactly know where the provider's resources are located.
4- Rapid elasticity: Computing resources can be elasticity obtained and released based on the current demand of consumer's needs. However, this key feature can significantly lower the operation cost.
5- Measured services: Resource usage can be controlled and optimised by using charge-per-use basis. This meter capability provides the transparency of the utilised service for both the cloud provider and consumer. Some companies like VKernal offer software to help the consumer analyse and cut down the cost of unnecessary resource consumption.

2.2 Classification According to Service Models

There are three primarily service layers according to the nature of service model used by the cloud providers. These three layers are essential business models which are normally cloud service provider sell to cloud consumer (Mell & Grance, 2011).

- **Infrastructure as a Service (IaaS):** The capabilities provisioned to the consumer are comprised of the computer hardware and software abstraction layer. Fundamental hardware resources are including processing power, storage and networks. On the top of this hardware is the software layer, called the hypervisor. The hypervisor task is to effectively decouple the operating system and its application from the underlying physical server (James, Shosha, & Gladyshev, 2012). Furthermore, the management stack at this layer which allows resource pooling in which many new servers can be added to the resource pooling achieving the concept/illusion of an infinite amount of resources (James et al., 2012). However, the consumer has full control over the operating systems, storage, deployed applications and has limited control over selected network components such as host firewalls (Mell & Grance, 2011). Amazon EC2 is a common example of IaaS. Users access virtual machines (VM) that are running on provider servers and can install any operating system and run any application on that VM. Furthermore, an image of the instance can be created by the user with the aim of saving the VM status and retrieving it later (Zawoad & Hasan, 2013a).
- **Platform as a Service (PaaS):** The platform layer is relying on the infrastructure layer. A web-accessible platform is provided for customers. The customer can acquire applications, test code and experiment with new software using program languages, libraries, and tools supported by the cloud provider. PaaS provides services to many concurrent users, these services are including storage, application development, and

hosting. Google Apple Engine (GAE) and Microsoft's Azure are two examples of PaaS in which developers can host their own developed web application on them (Zawoad & Hasan, 2013a). However, the consumer can control the deployed applications and configure the application hosting environment but has no control over the infrastructure layer this platform is built upon (James et al., 2012).

- **Software as a Service (SaaS):** The consumer uses the provider's applications which are run and hosted by a provider. The customer can access the application through web browsers and usually, there is a monthly fee for using this services (Zawoad & Hasan, 2013a). This layer is regarded as a top tier of cloud services architecture. It relies on the infrastructure layer in allocating more or less resources for the application's needs and relies on the platform layer in hosting services. However, the consumer has no control over the platform and underlying cloud infrastructure that the application is running on, or even on the application with possible limited access control over specific application configuration settings (Mell & Grance, 2011). Google docs and Apple's iCloud are well-known examples of software as a service (Yu & Wang, 2011).

There are also other service models which rely and are built upon some or all of the three primary layers aforementioned. Four examples are listed as the following:

- **Recovery as a service (RaaS):** This layer is required in case of failure or emergency; it essentially works as a backup of data to a cloud and makes it live by providing standby computing capacity on demand to facilitate more rapid application recovery. RaaS was defined by (Gartner, 2012) as "the managed replication of virtual machines (VMs) and production data in a service provider's cloud, together with the means to activate the VMs to support either recovery testing or actual recovery operations."

- **Security as a service (SecaaS):** Providing this layer of cloud services will enable enterprises to have more control over the security in the way that would not be cost effective if provisioned locally (James et al., 2012).
- **Database as a Service (DaaS):** Facilities are offered by the cloud provider. DaaS on the cloud utilises a multi-tenant architecture in which the same physical table can contain multi-user data. Most of the providers offer the customers abstractions in which users can store data in a key- value pair fashion rather than using the relational database. Amazon SimpleDB and Google Bigtable are two examples of DaaS's providers (Motahari-Nezhad, 2009).
- **Digital Forensics as a Service (DFAAS):** The aim of this layer is to provide digital forensics technology as a service. The centralised forensics labourites are built with forensics functionalities based on a large number of computing resources and securely control access in order to provide the investigators with the more computational power to conduct forensics processes which otherwise is time-consuming (Lee & Un, 2012).

2.3 Classification According to Deployment Models

Organisations have to consider many issues when moving to the cloud environment. Some Service providers are interested in lowering operation cost while other focusing on reliability and security (Zhang et al., 2010). Subsequently, there are different types of cloud as the following:

- **Public Cloud:** the cloud providers offer the cloud infrastructure as a service for open use by the general people or large industry groups. It exists on the premises of the cloud provider and can be owned and managed by a business, academics or government organisation. Cloud providers do have to invest initial capital on infrastructure.

However, public cloud has the lack of control over data, network, and security. Thus its effectiveness in many business scenarios can be hindered (Zhang et al., 2010).

- **Private cloud:** it also called the internal cloud, in which infrastructure is provisioned executively for a single organisation. It may exist on or off premises and can be owned and managed by the organisation itself, an external provider, or by a combination of both. Although, it is criticised as a capital investment on infrastructure has to be initialised, it offers the highest degree of the performance, reliability, and security (Zhang et al., 2010).
- **Hybrid Cloud:** it is the combination of public and private cloud models. In a hybrid cloud, most of the services can be deployed in a private cloud and the remaining will be run by the public cloud. The best split between public and private has to be carefully determined. In the case of peak demands, the non-critical application can be busted into a public cloud in order to meet peak demand and this concept is known as bursting cloud. However, hyper is more flexible than private and public clouds and have more control over application data. Hybrid cloud allows organisations to take advantage of the scalability and cost-effectiveness of the public cloud without risking sensitive data or mission-critical applications to third parties. Therefore, Gartner predicted hybrid cloud is going to gain in popularity and by 2017 it is penetrating 50% of large enterprises (ComputerWeekly.com, 2013).
- **Community Cloud:** when more than one organisation has the same concern such as mission, security requirements or policy. Accordingly, they can share the same infrastructure to achieve their shared concern or goals. This shared model is known as community cloud (Zawoad & Hasan, 2013a). Figure 2-1 illustrates the main service models the main deployment models and essential characters of cloud computing.

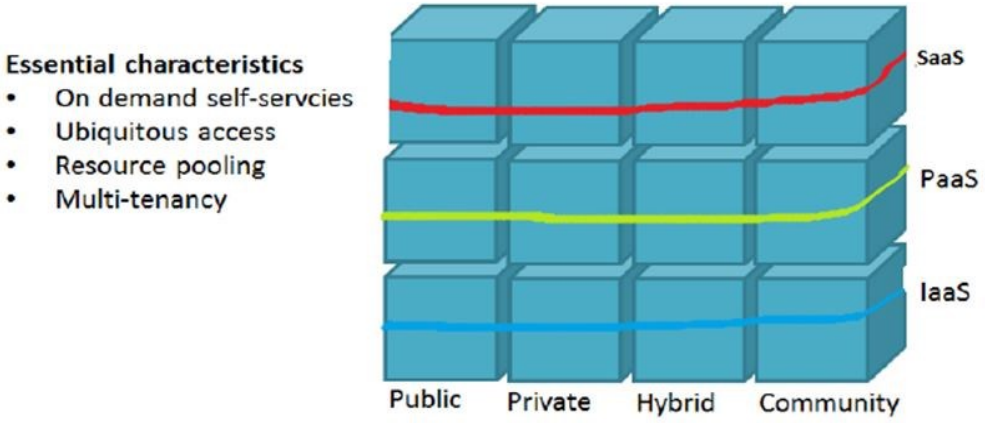


Figure 2-1 Essential Characteristics, Service Models and Deployment Models of Cloud (Almulla et al., 2013)

2.4 Cloud Computing Architecture

Cloud computing systems consist of two sections, namely the front end the back end. The front end is the side the computer user sees. The back end is where the system resides. They are both connected to each other via a network, usually the Internet. The front end contains the client's computer and applications needed to have access the cloud and the back end contains various computers, services and data storage. A central server called the middleware is responsible for administrating the system, monitoring the traffic and allowing networked computers to communicate with each other (Jadeja, 2012).

The architecture can also be expressed as a layer hierarchy consisting of the hardware/data centre layer, the infrastructure layer, the platform layer and the application layer, as shown in Figure 2-2 below.

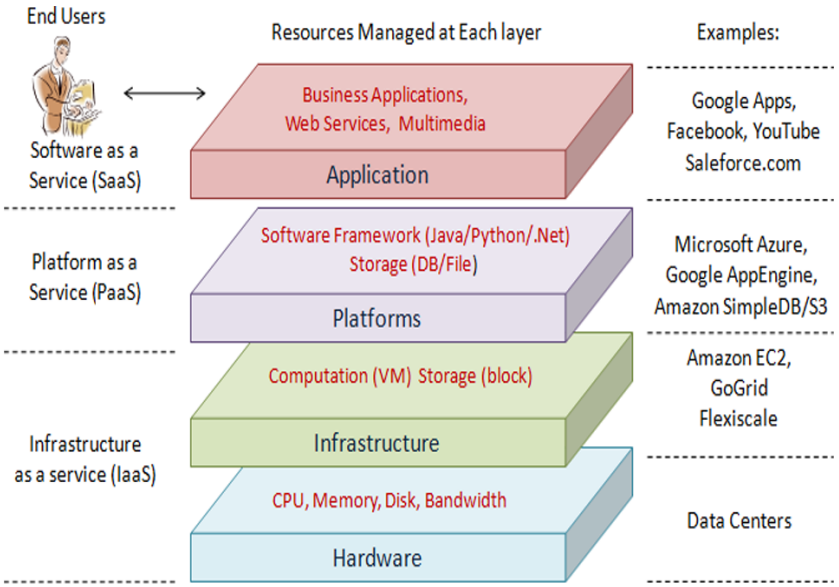


Figure 2-2 Cloud Computing Architecture (Zhang et al., 2010)

The main goal of the hardware layer is to manage the physical resources of the cloud which are typically implanted in data centres. A data centre usually has many of servers which are interconnected via switches, routers. However, there are typical issues associated with the hardware layer including hardware configuration, fault-tolerance, traffic management and cooling system management (Zhang et al., 2010).

In the infrastructure layer/virtualisation layer, a large pool of computing resources is created and partitioned by using virtualisation technology such as Xen, KFM and VMware. Virtualisation is a very important component of cloud architecture as it enables many key features such as dynamic resource assignment. The platform layer locates at the top of the previous layer (virtualisation layer). It consists of operating systems and application frameworks. The main purpose of this layer is to reduce the burden of deploying applications into VM containers. Ultimately, the Application layer which locates at the top of the hierarchy. It contains the actual cloud applications which have the automatic scaling features at aiming of getting better performance, availability and at very reasonable operation cost.

2.5 Related Technologies to Cloud Computing

There are certain technologies that are often compared to cloud computing because they share some technical aspects. These technologies are presented in this section, namely Grid computing, utility computing, virtualisation and autonomic computing.

Grid computing is where more than one computer coordinates and acts together with the aim of solving humongous tasks such as IT resource-intensive problems that otherwise handling them alone were too large and too complex (Adolph, 2009). Grid computing is used for problems involving a lot of number crunching in which can be easily parallelisable (Patel, 2014). Grid computing started from scientific and academic societies and spread over the commercial community. For example, HSBC banking group has employed more than 3500 CPUs which is running in different data centres in four countries to make numerous calculations and risk analysis based on available information and future events (Adolph, 2009). Another example is the German shipyard Flensburger Schiffbau Gesellschaft (FSG) which has created individual ship designs in a time and cost-effective manner by utilising high-performance computing resources in order to solve complex and intensive calculations. Having access to high-performance computing resources on demand, reduces the cost of ownership, reduces technical and financial risks in the ship design. The head of the basic design department at FSG, Thomas Gosch, claimed that "On-demand computing enabled us to solve CPU-intensive calculations in a very short time without the need to invest in an expensive IT infrastructure and specialized software at the yard. It helps us significantly in improving individual ship designs" (Dan, 2009). Grid Computing and cloud have the similarity in utilising distributed resources to achieve common objectives. Nevertheless, cloud computing has one more advantage by applying virtualisation technology to obtain dynamic resource provisioning (Zhang et al., 2010). Furthermore, a resource allocation of a task in cloud differs from

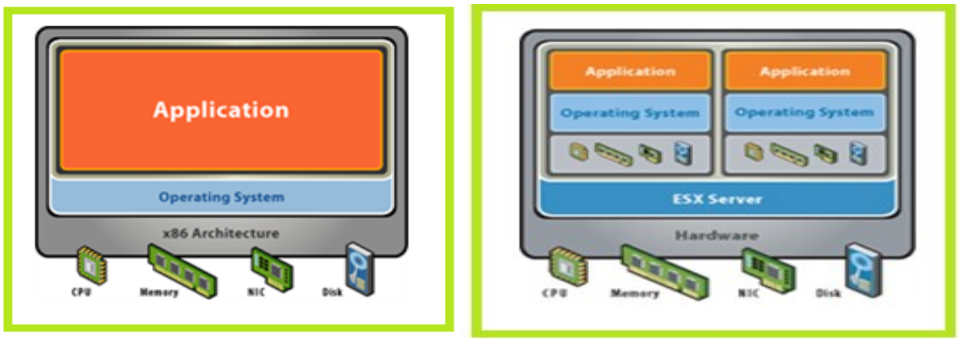
distributed grid systems. In the cloud, customers allocate a small fraction of the entire resource pool, allowing another customer to utilise separate fractions of the pool at the same time. Whereas, in the grid systems, customer allocate most, if not all, of the resource pool until the task is completed which means other customers have to wait till task before is complete (James et al., 2012).

The concept of utility computing is to rent an external utility computing resources including hardware, software, and network bandwidth on a demand basis rather than building servers in-house. Furthermore, utility computing allows customers to have computing resources on a demand basis and pay only for the used (Sourya, 2011). Having such features, services providers can increase resources utilisation and also decrease their operational costs (Zhang et al., 2010). Utility Computing delivers these resources to customers and gives them details about the instances, the volumes network services, having a label for a real VM instance on real hardware. To contrast this, a Cloud Computing provider would give you a flawless system, and the actual VM instance or real hardware behind it would forever be abstracted. The machine's configuration process is initiated with the knowledge that this machine is part of a pool (shared) (Red Eye Monitor, 2011).

Virtualisation abstracts the computer resources to provide virtualised resources, typically as virtual machines (VMs). In spite of the fact that virtualisation is not necessary to create a cloud, it is regarded as a key component of cloud computing. It enables rapid scaling of resources in a way that a non-virilised methodology finds it very difficult (Intel, 2013). However, Singh (2004) has defined virtualisation technology as *" a framework or methodology of dividing the resources of a computer into multiple execution environments, by applying one or more concepts or technologies such as hardware and software partitioning, time-sharing, partial or complete machine simulation, emulation, quality of service, and many others*" (Singh, 2004).

Despite the fact that virtualisation and clouds are technically different, they are very much intertwined. This section identifies the virtualisation process, types of virtualisation and the part of virtualisation takes place in a cloud environment. However, not all virtualised environments are a cloud but most cloud computing systems have virtualised components including servers and routers (Barrett, 2013). It is crucial to understand the difference between a physical machine and a virtual machine in order to begin the discussion around the virtualisation. A physical machine is the hardware devices such as personal computers or physical server which operates with direct access to the hardware whereas in the virtual environment the user can access data based on the name of the virtual machine file. Looking at the architecture of a physical server, it can be said that each server has its own hardware includes memory, network processing, and storage resources. On top of these hardware resources, the operating system is loaded and the application is running on the top of the OS, as shown in Figure 2-3a. By contrast, virtual machine architecture has the same physical server with all resources. The only difference is that instead of the server operating system, the virtual machine has a hypervisor where it can create the virtual machines (Academy, 2012). As shown in Figure 2-3b, each virtual machine has its virtual resources include virtual memory, virtual network, and its own virtual disk. On top of this virtual machine, a guest operating system is loaded and runs the application. The advantages of the virtualisation are becoming widely understood as it has the ability to run several guest operating systems and handful of applications running on a single physical computer if the computer has sufficient resources include processing, memory, and storage. According to this infrastructure, every single VM has its own Virtual hardware and the guest operating system is only aware of this hardware configuration. Thus, a VM is completely hardware isolated and independent and this makes the VMs are easily movable/portable from one physical machine to another. However, the hypervisor or Virtual Machine Monitor (VMM) is the part that responsible for providing and

managing several operating systems to run on the same physical machine at the same time. There are different type o VMM such as bare-metal, hardware assist and hosted (Barrett, 2013).



(a) (b)

Figure 2-3 Physical Machine vs. Virtual Machine (Academy, 2012)

Besides being cost effective by decreasing the number of physical machines, many benefits are offered by virtualisations including being more effective in disaster recovery, better resource management such as server consolidation and testing, isolating and testing malware actions (InfoWorld, 2011). However, there are drawbacks of using virtualisation including higher complexity to manage - in the case of intrusion, an intruder can gain control of all the guest operating systems and the risk to the sensitive information is greatly increased. For example, the Blue Pill attack that controls the interaction between the host (hypervisor) and guest (virtual machine) (Kevin, 2006). Thus more consideration for the data must be taken into account (Barrett, 2013). It is vital to understand the relationship between virtualisation and the cloud. The biggest difference between them is that virilisation is technology based whereas the cloud is model based. Furthermore, virtualisation is one physical computer that tries to be many computing environments whereas cloud works in the exact reverse, many different computers try playacting to be one computer environment. Also, the purpose of virtualisation is to maximise machine resources whilst the cloud purpose is to provide the ability for business to have access to resources without maintaining the physical resources. Cloud computing can

deliver services while virtualisation can be one service available for delivery. However, many cloud providers use virtualisation as the basic components for cloud due to its flexibility and ease of VM movement. Thus, virtualisation is usually a counterpart for cloud environment and it is a subset of the service requirement for meeting the needs of organisations (Phillippi, 2010).

Autonomic Computing has been coined by IBM in 2001. The idea is to build the computing systems that have self-management built in feature in order to react to any observation without needing to human interaction. Cloud computing provides some autonomic features such as automatic resource provisioning. However, cloud computing aims to reduce the cost of resource whereas the main goal of autonomic computing is to reduce the system complexity (Zhang et al., 2010). In summary, comparing utility, Grid Computing, autonomic computing to cloud computing, it can be said that cloud computing does everything they do and much more (Sourya, 2011).

2.6 Summary

Cloud computing has recently emerged as a compelling and new paradigm for hosting and managing service over the internet. Cloud computing has changed the landscape of information technology and turned the long-held promises of computing services into reality. This chapter covers essential concepts of cloud computing including its definition, the prominent and essential characteristics and the most common cloud computing service models (i.e. IaaS, PaaS and SaaS). It also introduces the main deployment models including Public, Private, Hyper and Community cloud along with identifying benefits and downsides for each deployment model. This chapter shows the basic service components of cloud infrastructure and presents studies which prove that cloud computing technology is important to the industry. This chapter also highlights the unique combination of characteristics that that impact upon security and digital forensics including; on-demand self-service, broad network access, resource pooling, rapid

elasticity and measured service. Ultimately, it discusses related technologies to cloud computing and its architecture.

3 Digital Forensics

3.1 Introduction

There has been a sharp increase in Internet users, with Internet usage statistics showing that from 2000 to 2016 there is an estimated 825% increase in the number of Internet users worldwide, from 414 million in 2000 to more than 3.5 billion at 2016 (Internet Live State, 2016). Furthermore, Business Insider (2015) stated that the world has 10 billion devices connected to the Internet in 2015. Moreover, Ericsson CEO, Hans Vestberg, predicts that by 2020 these connected devices will increase to 50 billion and incorporate non-traditional devices such as TVs, kitchen and house appliances (Higginbotham, 2015). Clearly, the trend of Internet user and Internet-connected devices is increasing at a rapid rate. Unfortunately, this rapid evolution of technology has caused these devices to be used in criminal activities. Thus, the number of digital related crimes is sharply increasing too (Koen, 2009, Trenwith & Venter, 2013). Cybercrime is a growing trend around the world. Statistics show that the number of victims of cybercrime is around 556 million per year, equating to 1.5 million per day, or 18 per second (InSecPro, 2014). Furthermore, cybercrime activities cause an annual loss of 100 billion dollar to U.S companies alone, which is roughly equivalent to 500,000 jobs (Lewelling, 2013). Subsequently, the amount of data that must be examined in a digital forensics continues to rise at a very high rate due to the complex and dynamic developments associated with the Internet, digital Information and evolution in communications technology (Miller et al., 2014). However, the rapid advancement of devices and users connected to the Internet makes the situation even more difficult for digital forensics. Based on the fifth annual cost of cybercrime study conducted by the respected Ponemon institute on behalf HP Enterprise Security, the average time to resolve a cyber-attack had significantly increased, climbing from 32 days in 2013 to 45 days in 2014 (Ponemon-Institute,2014). Furthermore, the average annualised cost of cybercrime has reached $12.7 million for companies in the US; a 9% increase comparing to

11.6 million in 2013 and a 96% increase since the study was launched five years ago (eSecurity, 2014). Hence, it is imperative for forensic investigators to keep up to date with these developments and to identify areas in which research and development are needed. It is crucial to facilitate and improve the efficiency and effectiveness of the digital investigation by enhancing tools, techniques and scientifically proven methods that used when carrying out the investigation process.

3.2 Defining Digital Forensics

In the 1980s, law enforcement was not able to cope with computer crime based investigations. The way to examine the digital evidence is completely different from traditional evidence as the examiner deals with non-physical evidence which cannot be observed without interpretation. Thus, the need for a specialised team to handle such investigations was realised. With growing usage of computers in the 1990s, computer forensics was developed as an independent field (Evann et.al, 2011). However, in the early days, evidence collected was confined to computers, thus it was called computer forensics. Due to the proliferation of more advanced technological devices, a new term called digital forensics was used instead of computer forensics to act as an umbrella for all digital-based forensic investigations.

The term "forensics" is the key word that should be understood before defining digital forensics. The term forensics is a scientific methods or techniques used in connection with the detection of crime and relating to the court of law (Oxford Dictionary, 2014). Despite a multitude of technological and philosophical changes to the field of digital investigation, the definition proposed by the first Digital Forensics Research Workshop (DFRWS) is still a widely accepted and remained popular since the time of this definition (James et al., 2012). Carrier (2001) stated that the digital forensics science was defined by DFRWS as " *the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis,*

interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations".

Digital evidence is any data that can provide a significant link between the perpetrator and the victim (Wang, 2007). Such evidence must be retrieved in a forensically sound manner that will be accepted in the court. However, digital evidence has its unique physical characteristics that pose new challenges to digital investigators. These characteristics were identified by (Casey, 2001) as follows:

- Digital evidence is easy to be copied and modified, but it is not easily kept in its original state. As the digital form stored in a computer system is binary 0 or 1. The copied object is exactly the same as the original, but it is also easily to be amended with user modifications. In turn, it is difficult to retain digital evidence in its original status and therefore proving of the original digital source is very challenging.
- It is difficult to prove the integrity of digital evidence. Establishing a direct link between the evidence and the suspects is very difficult because it is very easy for the user to copy and modify the evidence.
- Human senses cannot observe the presentation of digital information unless there is help of a suitable toolkit.

With each advancement in technology the forensic investigator experiences new challenges (Raghavan, 2012). However, the field of digital forensic has been continually developed and several process models have been proposed. There is no one standard developed with global acceptance (James et al., 2012). Therefore, the majority of organisations create their own standard which is not directly based on common standards.

3.3 The Process of Digital Forensics

The ultimate goal of the digital forensic process is to produce digital evidence that could be admissible in the court of laws (Law & Chow, 2014). The majority of the literature studies agree on the activities that facilitate the digital forensic investigation process. It encompasses different domains of computer science, legal concepts and principles as well as the steps of identifying, acquiring and preserving, examining, analysing, documenting and reporting the evidence by an expert witness in a court of law (Raghavan, 2012). Figure 3-1 shows the sequence of these stages.

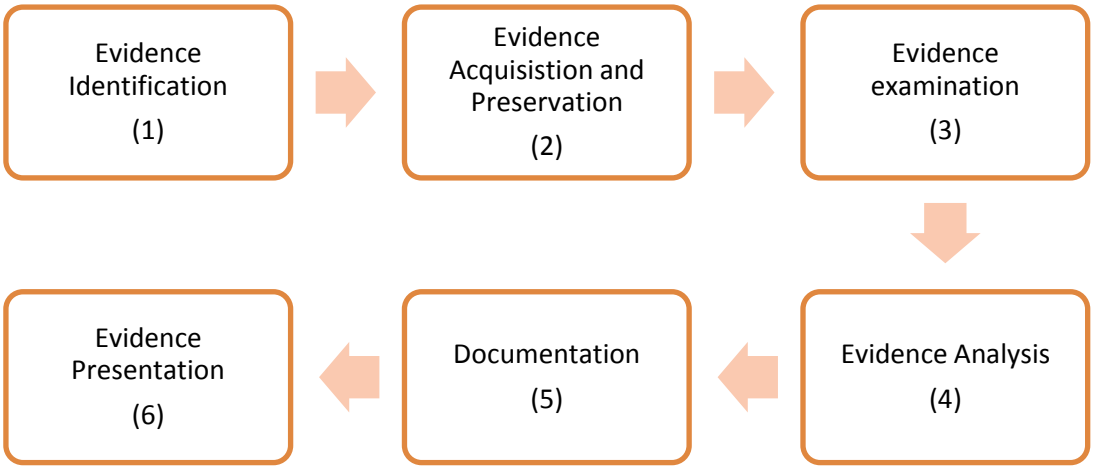


Figure 3-1 Main Stages of Digital Forensic Process

The following commentary identifies and discusses the main steps performed during the process of digital investigation.

1- Evidence Identification

In this stage, there are two things have to be identified by digital investigators. First, the potential sources of data that can provide digital evidence including, but not limited to, CDs, DVDs, zip disks, hard disks on computer systems, random access memory

cards, PDAs, mobile phones and so on. Second, the relevant, admissible data which is stored on these sources (Fei, 2007).

2- Evidence acquisition and preservation

Once the sources and its data are identified, the task of acquisition begins. At this task, digital evidence has to be acquired and forensically preserved. To do so, forensic evidence identified by a forensic people need to be imaged by obtaining a binary bitwise copy in order to avoid tampering with the original copy of the evidence. Moreover, it is imperative at this stage to protect the integrity of all evidence. The universal tool used to verify the integrity is the Hash Function. The common standard hash functions are MD5 or SHA-1,SHA-2 and SHA-256 (Pearson, 2014). Furthermore, it is crucial to make multi-read only copies and conducting all forensics tests. Thus the original copy kept safe and any tampering with the original copy of the evidence is avoided (DFRWS, 2006).

3- Evidence Examination

In this step, forensics investigators are using forensics tools, filtering and data extracting techniques in order to identify artifact of interests and make them ready for the analysis stage. Furthermore, this includes extracting relevant implicit and hidden information which is called data recovery (Ademu et al., 2011).

4- Evidence Analysis

This stage refers to conducting the application of scientific methods in order to answer the essential questions in an investigation: what, who, why, how, when and where (Casey, 2011). Investigators in this stage attempt to reconstruct the sequence of events by analysing the data collected during the previous stages. The data is analysed thoroughly with aim of discovering any traces of evidence which could be useful from the technical as well as the legal perspective (Ieong, 2006). Indeed, finding the

relationship between the evidence and the incident, creating the sequence of events chronologically and linking together artifacts helps investigators to understand the complete picture of the case (Yadav, 2011). Analysis can be performed in one of two modes; dead or live. When the system is in running states, then it called live system and analysis at this situation is known as live analysis. Once the machine is shutdown or the plug is removed, all volatile data will be lost and the analysis is dead. Thus, the collection of volatile information is more important for forensic analysis (Yadav, 2011). It is recommended to collect the volatile data such as system hardware information along with the TCP and UDP ports are open, user login history, activated services, etc. (Hunt & Zeadally, 2012). Researchers have developed script programs to collect such volatile data to the forensics results at aiming of reducing barriers to operating and facilitate analysis (Yen et al., 2009). Dead forensics analysis, however, conducted when the system is not in running states.

It is not easy to provide standard operating procedures to cover all aspects in depth of forensics analysis at the digital scene. This is due to the difference between the suspicion nature of the case/incident and digital crime scene situation. Thus, it is crucial to create a methodology/approach that organises and analyses such huge volume of data extracted from digital devices including computers, laptops and networks. In general, crime reconstruction provides such a methodology (Valjarevic & Venter, 2013). The goal of crime reconstruction is to exploit knowledge of the series of events that surround the commission of a crime using deductive reasoning, physical evidence, scientific methods and their interrelationships (Knox, 2012). In crime reconstruction, there are three approaches which can be performed in order to reveal a crucial interaction as the following:

• Relational, where there is a relation between two or more objects and how they are connected to each other's.

• Functional, how some this works and how was it used.

• Temporal, refers to the time that event related to.

In 1940, a French criminologist and forensics science pioneer, Dr. Edmond Locard articulated one of the science's key rules, known as Locard's Exchange Principle. The principle states that every contact leaves a trace (Mohay et al.,2003). According to this principle, whenever criminals commit a crime, they leave some evidence at the scene which was not there before a crime has been committed and they take some item of evidence when they leave the scene. This principle of exchange is applying whenever there is contact between people or between objects and even between people and objects. Likewise in the digital crime scene, "when you go on the Internet, you leave fingerprint – we can tell where exactly you have been" Sheriff's investigator Mike Gurzi, said (Casey, 2001). However, Locard's exchange principle is the heart of crime scene reconstruction (Knox, 2012). The geographic location of people, computers, and communication should be included to create a full relational reconstruction. Moreover, creating the list of IP addresses connections and sorting them according to the destination can reveal the crucial interaction and draw the picture of how computers interact. Also, it is important that a forensics examiner determines how a particular system was running and configured at a particular time in order to gain a better understanding of a piece of digital evidence. Furthermore, establishing the time line of events and identifying their sequence of can help in discovering patterns, gaps and may lead to identify further related evidence (Casey, 2012). However, it is necessary to produce an accurate timeline; thus different time zones and system clock accuracies have to be taken into account.

5- Documentation

All critical details of each step of the digital investigation process have to be documented. This includes procedures performed and tools used to seize, collect and analyse the forensics data. However, this step is crucial to support the digital investigators in defending their hypothesis about the existence of particular evidence. Especially in the case of that the criminal claims the evidence was done by somebody else and repudiates doing anything wrong (Fei, 2007).

6- Evidence Presentation

Finally, findings are presented in a court of law and usually supplemented by an expert witness for testifying (Raghavan, 2012).

3.4 Forensic Standards

A Standard Operating Procedure (SOP) is a set of steps that must be followed when a computer is collected or examined. These steps are needed to grant the consistency and thoroughness and to ensure that the best available methods are used. One of the best and useful guides of handling computers as evidence is The Good Practices Guide for Computer-Based Evidence published by the Association of Chief Police Officers in the United Kingdom ACPO (ACPO, 2012). This guide provides useful guidelines, flowcharts and template forms for the initial examination of a computer. There are four principles are involved in this guide as the following:

Principle 1: No action taken by law enforcement agencies or their agents should change data held on a computer or storage media which may subsequently be relied upon in court.

Principle 2: In circumstances where a person finds it necessary to access original data held on a computer or on storage media, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.

Principle 3: An audit trail or another record of all processes applied to computer-based electronic evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.

Principle 4: The person in charge of the investigation is responsible for ensuring that the law and these principles are adhered to.

3.5 Digital Forensics Tools

General purpose forensics tools have greatly improved investigator's ability to analyse digital evidence and have made the investigation process effective and efficient; otherwise, the process of digital forensic will take a huge time and effort to be conducted in a forensically safe manner. Thus, each investigation will determine the tools that are used. Nevertheless, there are several commercial and open source digital forensic tools dedicated for the purpose of forensic investigation. It would not be an easy task to specify/label the best tool, technique or methodology compared to the rest as most of these tools have their advantages and disadvantages. Paraben Corporation which is specialising in the developing computer and mobile phone forensic tools produced the comparison reviews of a variety of today's forensics tools including AccessData FTK5, Guidance Software EnCase Forensic 7, X-Ways Forensic, Belkasoft Evidence Center, TechPathways ProDoscover and Paraben's P2 Commander 3.0 (Paraben, 2013). This comparison/report are compared against the capabilities of four technical points of interest: the capability to performing file analysis, the capability to search, the capability to perform forensic analysis and capability to report and export the results of forensic analysis. It is clear that FTK and Encase have advantages over the rest of the tools in dealing with a variety of file types in terms of conducting file analysis where as TechPathways ProDoscover tool cannot analysis majority of file types such as Windows Mail E-mail, E-mail Files (EML), Skype Chat and Safari Browser History. It is clearly visible that Paraben's P2,

FTK and EnCase can perform the major search capabilities including Hex Searches, Boolean Searches and Indexed file searches whereas Belkasoft Evidence Center is less than other tools in performing searching functions. Encase and FTK are also the most capable tools in conducting forensic analysis functions including physical disk & disk image browsing, unallocated space browsing, deleted file recovery, file slack browsing, detecting file signatures, calculating file hash codes etc. Ultimately, in terms of reporting and exporting the result of the forensic analysis, Paraben's P2, FTK and EnCase can report and export examination results in different formats such as XML, HTML, CVC, Text report, E-mail Attachments Export etc. However, TechPathways ProDoscover and X-Ways tools can only support one or two of these formats.

Overall, Encase and FTK tools have the most capabilities and features to conduct major computer forensic functions. Despite this, Ayers (2009) points out that they do not support investigation at a sufficient level and as such, they have been described as the first generation forensics tools that have limitations. These limitations include; reliability, audibility, data abstraction, efficient data storage and repeatability of forensics analysis results. The author suggests that the need for developing the second generation of tools has become apparent in order to address the shortcomings of the currents 1st generation. Furthermore, the author lists set of requirement for the 2nd generation were introduced namely: parallel processing, high performance and scalable data storage, accuracy and reliability must be 100% at all validation tests, auditability that maintains an audit trail record of all the analyst's activities, repeatability and data abstraction should be at a high level for various types of documents such as generic text, email messages, audio and video files (Ayers, 2009).

On the other hand, open source digital forensic tools are assessable solving crimes. For example, the Sleuth Kit®, Autopsy®, Raptor and other open source digital investigation tools.

Furthermore, open source tools are more user-friendly and much easier than commercial forensic tools. Its capabilities are including data acquisitions and analysis. However, there is a drawback of utilising open source tools. These drawbacks are including ambiguity in the requirements, the complexity of the software for less technical people and absence of standards (ARUN & GANESH, 2005).

3.6 Additional Issues

In certain aspects of the digital investigation, many challenges are faced by digital forensic investigators. For instance, gathering data from areas of the hard drive that are beyond the reach of most users, for example, unallocated space, file and disk slack space, and so on (Lillard, 2010). Povar & Bhadran (2010) pinpointed that recovering files from digital media for which no file system information is available is a critical problem that faces digital investigators. Several reasons make the file system information is unavailable. These reasons are including formatting of the digital media, the file of interest may have been deleted such that the file system indexes no longer refer to the file content, a file can be hidden in areas like lost clusters, unallocated clusters and slack spaces. As such, such items may require special processing and some degree of independent preservation, survey, and examination in order to extract usable information from them (Eoughan Casey & Schatz, 2011).

Indeed, deleted data or partial file data helps an investigation which gave rise to the new field of data carving (Raghavan, 2012). Data carving is an essential aspect of computer forensics and is an area that has been somewhat neglected in the development of new forensic tools (Alherbawi, Shukur, & Sulaiman, 2013). Digital Forensic Research Workshop (2001) defined data carving as the process of extracting a collection of data from a larger data set. Data carving techniques frequently occur during a digital investigation when the unallocated file system

space is analysed to extract files. The files are carved from the unallocated space using file type-specific header and footer values.

Furthermore, the increasing volume of data and time taken to perform the investigation process are still obstacles. Roussev (2009) pinpointed that media capacity continues to double every two years. Thus, huge data sets will increasingly be the norm. Therefore, accurate and efficient automation techniques are highly needed. Al Fahdi et al. (2013) placed the emphasis on the need to develop approaches that identify and extracts "significant data" through techniques such as criminal profiling is becoming as essential privacy issues are other issues that need more attention and research. Privacy issue arises when investigators can breach the secrecy of unrelated data. For example, utilising knowledge that out of the scope of the investigation such as disclosing private images, encrypted key and the user passwords (Aminnezhad & Dehghantanha, 2012). Moreover, the jurisdiction of data is one of the most data challenging topics which seem to be overlooked (Damshenas et al., 2014).

These issues and others have become even more complicated when massive shared infrastructure such as cloud computing is involved. There is a lack of forensic tools that are tailored for a cloud system. According to a survey conducted by Shah et al. (2013), approximately 58% of respondents agreed that digital forensic process automation is needed to tackle future challenges including cloud forensics. Additionally, there is a high level of demand on the forensic-aware tools for the CSP and the clients to conduct a forensic investigation in the cloud environment. Hence, it is crucial to develop tools which can be utilised to identify, collect, and analyse cloud forensic data. Therefore, and in order to assess the changes that will need to be made to undertaken digital investigations in a cloud environment, an in-depth analysis of cloud forensic issues is conducted in the next chapter.

3.7 Conclusion

Statistics presented in this chapter show that the trend of Internet users and Internet-connected devices is increasing at a rapid rate, this rapid evolution of technology has caused these devices to be used in criminal activities and therefore the number of digital related crimes is also sharply increasing.

Fortunately, commercial and open source digital forensic tools including Access Data (FTK), Guidance Software (EnCase), X-Ways have greatly improved the investigator's ability to conduct major forensic functions. However, they do not support investigation at a sufficient level as they still have limitations including reliability, auditability and repeatability of forensics analysis results.

There are many different challenges exist within digital forensics. These issues and others have become even more complicated when massive shared infrastructure such as cloud computing is involved as the current forensic tools do not have cloud forensics capabilities. The analysis shows that the available forensic tools have various limitations and cannot cope with the distributed and elastic characteristics of cloud computing. Therefore, further development of current tools has become necessary in order to address their pitfalls.

4 Cloud Forensics

This chapter aims to establish the knowledge of key fundamentals of cloud forensics including its definition, current research proposals and technical solutions addressed with the aim seeking to evaluate the extent to which current forensics techniques can be applied. Furthermore, it aims to highlight the open problems that need further effort to be tackled.

4.1 Introduction

Cloud forensics refers to conducting all the processes of digital forensics in the cloud environment. It is a cross-discipline concept between digital forensics and cloud computing. However, some researchers such as Ruan et al., (2011), defined cloud forensics as a subset of network forensics – which is an investigation technique looking at the network traffic generated by a system for the purposes of information gathering, legal evidence, or intrusion detection (Cohen, 2008) – due to the heavy reliability on extensive network access alongside the fact that network forensics handles forensics investigations in private and public networks. Despite the many benefits cloud computing introduces for organisations it also presents opportunities for criminal exploitation leaving almost no evidence behind. Indeed, the issue of security is listed as the top concern of cloud adoption (Hooper et al., 2013). The distributed nature and configuration of cloud computing infrastructure creates a range of problems for digital investigators. Whilst the question of whether cloud computing systems have the ability to support and perform digital investigations in a forensically sound manner has been stated previously, a significant lack of research exists (Ruan et al., 2013). The current methodologies, procedures, tools and architectures are not designed to handle and assist the digital forensics in cloud environments; notwithstanding, the fact of on-going and proactive investigations are becoming mandatory components for enterprises (Poisel et al., 2012). Therefore, solutions that provide cloud forensics must be sought. To date, researchers have focused on identification of

the foremost issues facing digital forensics investigators when performing digital investigation in cloud computing. Investigators have limited success to keep up with the currentt development of cloud when applying traditional forensics investigation approaches and tools. For example, memory copy and evidence storage devices are not physically available anymore unlike non-cloud environments – all cloud environments are virtualised and hardware resources are shared between the virtual machines (Aydin & Jacob, 2013). The reason behind this is that the evidence is not stored on the regular basic server infrastructure which was in hand and controlled by users. Furthermore, the original sources of evidence might be distributed in several virtual machines (CHEN, 2014). According to the survey conducted by Ruan et al., (2013), based on 156 forensics experts and practitioners located all over the world, more than 50% of the respondent agree that "establishment of a foundation of standards and policies for forensics that will evolve together with the technology" is an opportunity for cloud and 88.89% of them strongly agree that "designing forensics architecture for the Cloud" is a valuable research direction for cloud forensics.

4.2 Forensics Artefacts in Cloud Computing

Digital investigators in cloud-based crimes may investigate file systems, process, cache, and registry (Zawoad & Hasan, 2013b). However, the nature of the digital forensics investigation varies based on the service model of cloud computing: Software as a Service, Platform as a Service and Infrastructure as a Service. Three basic components were identified by NIST for delivering cloud services include Physical Resource Layer, Resource Abstraction Layer and Service layer as shown Figure 4-1 (Liu et al., 2011). A list of forensics artifacts was identified for the cloud system stack by Ruan & Carthy (2012a) as follows:

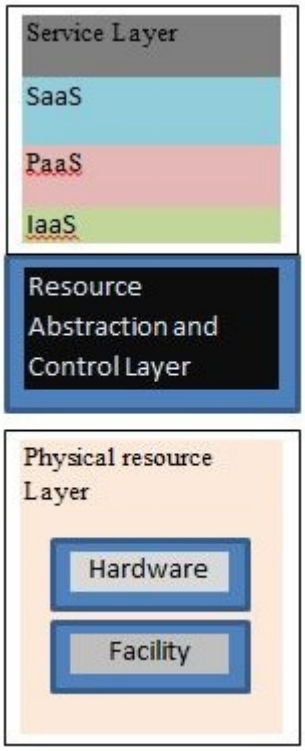


Figure 4-1 Cloud System Environment

- The Physical Resource Layer contains the physical computing infrastructure elements including computers (CPU and memory), Networks (all network devices such as routers, firewalls, switches network links), and storage elements such as hard disks. The physical resource layer is often geographically distant from both the cloud consumer and law enforcement and it is under control of the cloud service provider (CSP). Thus, to carry out on-site investigation, it is necessary to conduct a remote forensics or relying on the CSP to provide the required data. Networking is, however, one of the main enabling elements of cloud computing (IBM, 2012). In the cloud, there are two types of network. Firstly, the physical network which connects physical hardware such as nodes, network storage, and Internet connection. Secondly, the virtual network which responsible for connecting instances to each other and connect the virtual network to the physical network. However, from a digital forensics point of view, there is crucial

evidence that is sent over the physical network consisting at least of two computers (Fei, 2007). Investigators can capture and analysing network traffic in order to gather possible evidence. There are also other sources of network forensic evidence including logs from servers and router information. Network forensics, however, can be conducted live and common tools to collect such data include TCPDump, Wireshark, and TCPFlow. Nevertheless, performing live analysis will cost significant hardware resources on a network consisting of more nodes than a typical home network (Delport et al., 2011). The facility resources such as air conditioning, power, and ventilation are also in the physical resource layer.

- The Resource Abstraction and Control Layer consists of the system components that enable cloud providers to manage access to the physical component resources via software abstraction. These components include hypervisors, virtual machines and virtual data storage (Liu et al., 2011). This layer is controlled only by the cloud provider and distant from the cloud consumer. However, this layer plays the critical role in solving unique cloud computing issues including a segregation evidence for multi-tenant and locating the actual physical resources from virtual resources in the Service Layer. However, the virtual images and hypervisor event logs are the important forensics components in this layer.
- In the Service Layer, the cloud consumer can use interfaces to access the computing services of each of the three service models that are provided by the cloud providers. As shown in Figure 6, SaaS applications can be built on top of a PaaS platform and PaaS can be built on top of an IaaS platform (Liu et al., 2011). The following paragraphs will briefly discuss each platform and illustrate the important forensic components in each layer.

- OS Layer (IaaS): this layer can provide an IaaS consumer with access to an operating system and drivers and it is hidden from SaaS consumers as well as PaaS consumers. However, the IaaS consumer is fully responsible for the guest OS`s whereas the IaaS provider is fully responsible for the host OS. The important forensics components in this layer include virtual operating system event logs, configuration logs, audit logs, registry, anti-virus application logs, intrusion detection system logs and virtual network logs.
- Middle Ware Layer (PaaS): This layer can serve the developer of the application in cloud by providing the interface of software building blocks such as libraries, database and Java virtual machine. The most important components in this layer include source code, performance logs, debugging logs, access logs, and account information.
- Application Layer (SaaS): This layer includes software applications that are used by SaaS consumers and might be installed, managed and maintained by SaaS providers, PaaS consumers and IaaS consumers. The important elements in this layer are very similar to forensics components in traditional software applications except that the software is hosted remotely from the consumer. Consequently, data collection in this layer is totally different from traditional software applications. However, the most important components are including application logs, authentication and authorisation logs and account information.

4.3 Current Cloud Forensic Frameworks

NIST stated that there is a lack of guidance on how to acquire and conduct forensics in a cloud as it is problematic or impossible to seize the physical hard drive in a cloud environment (Mell & Grance, 2011). Furthermore, Birk & Wegener (2011) observed that the current best practice guides in gathering digital evidence from cloud are rare and outdated. Therefore, it is crucial

to develop new approaches in order to extract evidence from a cloud environment. Also, there is a lack of academic research to enhance and develop such techniques that collect digital evidence in a forensically sound manner (Huber et al., 2011). This observation, however, found an echo of many forensic practitioners who postulated that "More research is required in the cyber domain, especially in cloud computing, to identify and categorise the unique aspects of where and how digital evidence can be found. Furthermore, end points such as mobile devices add complexity to this domain. Trace evidence can be found on servers, switches, routers, cell phones, etc." (ZatykoDr & Bay, 2011). From the legal point of view, applying current standards for traditional server-based systems will create potential difficulties for the digital forensic analyst due to the difficulties in finding the specific location that data is stored and what software on which specific device was installed in a cloud environment (Taylor et al., 2011). To date, researchers in cloud forensics focus on the three areas, cloud forensics frameworks, data acquisitions, issues and challenges (Thethi & Keane, 2014).

There are, however, two widely used frameworks in the digital forensics area, (McKemmish, 1999) and NIST (Kent et al., 2006). Upon these two frameworks, an integrated iterative conceptual cloud framework was proposed by (Martini & Choo, 2012), to provide a better support for conducting a digital investigation in a cloud environment. The next section addresses each framework focussing on the integrated iterative conceptual framework.

McKemmish (1999) has proposed a framework that encompasses four steps as the following:

1- The identification of digital evidence: identifying what evidence is present, where and how is it stored for the purpose of facilitating suitable techniques to recover it. This process can extend to any digital device that can be used for storing digital data including mobile/cellular telephones, electronic organisers (digital diaries) and smart cards.

2- The preservation of digital evidence: to ensure that any examination of digital data was conducted in a way that the least amount of changes to data occurs. In the case where changes are unavoidable, it is crucial to explain the nature of changes and the reason for the change. This applies to the changes made to the data itself and to the devices used to access the data.
3- The analysis of digital evidence: to extract, process and interpret the preserved evidence in a humanly meaningful manner.
4- The presentation of digital evidence: to present the results in a court of law.

Kent et al., (2006) from NIST has provided a Guide to Integrating Forensics Techniques into Incident Response including that conducted when performing the digital investigation process. The NIST framework consists of the following steps:

1- Collection: to identify, label and acquire data from the possible resources.
2- Examination: using forensically methods to process, assess and extract data while maintaining the integrity of the data.
3- Analysis: where the results of the examination are analysed using admissible methods in order to address the questions that arise the reason for conducting the examination and collection step.
4- Reporting: reporting the results which derived from the analysis step.

However, the process of digital evidence that involves cloud-based evidence should be viewed as a cyclical process (Quick & Choo, 2013). An examiner should not stop analysis of evidence already seized and wait for evidence identified and provided from the cloud environment. The examiner should analyse the evidence in hand and once the cloud data is provided, it should be included under the scope of the investigation for analysis.

Martini & Choo (2012), have proposed a cloud-specific an integrated iterative conceptual framework. Their framework combines the two widely used frameworks of (McKemmish, 1999) and NIST (Kent et al., 2006) which have been addressed. It seeks to provide better support to digital forensics investigations in a cloud computing environment. However, there are two key differences that clearly appear on Matrin and Choo framework:

a- Due to the nature of data in a cloud environment such as its physical location, distribution across multiple data storage and multi-jurisdictions. Thus, extraction of evidence from the cloud is critical and consequently, the collection step is presented as a separate step named "collection". However, this stage was included as part of the analysis phase in the McKemmish's framework. The importance of this step has been noticed by other researchers such as (Birk & Wegener, 2011);

b- The iteration phase which gives the investigators the ability to identify and preserve the evidence during any stage of the investigation, even after it is discovered in the examination and analysis step. For example, if an examiner found the evidence in the examination and analysis step, simultaneously, iteration will commence with evidence source identification and preservation.

However, the framework of Matrin and Choo comprises the following four steps:

1- Evidence Source Identification and Preservation: this step is concerned with identifying possible evidence. However, the first iteration is likely to identify the evidence from a physical device such as PCs, laptops, tablets and mobile handsets. The second iteration will identify cloud providers and then possible evidence stored in the cloud. In all cases, preservation of the evidence has to be taken into account.

2- Collection: where actual data is collected. In the cloud, there is a variation in the way to collect data, as it will depend on the type of cloud deployment model. For example,

if the data persists on Infrastructure as a Service model then the virtual hard disk and memory can be exported, whereas if the platform is Software as a Service, then only a binary export of the hosted software can be exported (Martini & Choo, 2012).

3- Examination and Analysis: discovering cloud evidence during this stage can lead to a second or may be more iterations of the process as mentioned earlier.
4- Reporting and Presentation: report and present the findings to the court of law. However, this step does not differ from the frameworks of McKemmish and NIST.

4.4 Conducting Forensics Examination in the Cloud

The first question raised here is who will conduct the exam and the second question is where this exam will take place. Possible choices for the first question are included law enforcement, cloud service provider's technician or independent third party examiner. The possible choices for the second question are included providers headquarter, the remote data centre of the CSP, a remote law enforcement lab or at the third-party lab. A cloud investigator can determine what layer of the cloud to execute the forensics process. However, there are six layers of the cloud that forensics process can be executed on. These layers are identified by (Dykstra & Sherman, 2012) as the following:

1- Physical Hardware
2- Network layer
3- Host OS
4- Virtualisation
5- Guest OS
6- Guest Application/ Data.

Technical capability and the trust in the data returned have to be considered when conducting the forensics process at each level. Each layer has its type of data, for example, the Network layer has packet captures, the Physical Hardware has physical files and the Host OS contains virtual files. It is crucial to ensure that for each type of data at each layer of cloud, the chain of custody is maintained admissible mechanisms and integrity checking must be included. Furthermore, taking snapshots of a VM introduces one powerful new option in which examiner can take a snapshot of a running machine, restore it later at his/her convenience time, run it as if it was life and perform live forensics after the fact. Dykstra & Sherman (2012), conducted three experiments to examine the ability and scientific accuracy of using the most popular forensics tools to acquire cloud bases data. However, the first experiment was to collect data remotely from the guest OS (Layer 5). There are no necessary changes required to the cloud infrastructure and, nor assistant from the provider. However, there is an extra cost that associated with a remote forensics investigation due to the significant bandwidth which is required to image and retrieve a virtual hard drive, for example, amazon charges outbound data transfer at $0.15 per GB, for the first 10 TB/month. Then $150 is the cost for transferring one TB of data. The second experiment is to collect data from the virtualisation layer by the injection of EnCase Servlet or FTK Agent. The considerable change made by the provider to the cloud environment is necessary. The third experiment is to collect data from host OS by utilising Amazon's export feature. Amazon provides specific data by importing this data into portable storage devices and ships it back to the requestor. The requestor of the data has to provide the devices and charged per storage handled and per data loading hour (Amazon, 2014). However, data export is executed by the provider and does not require any changes to the infrastructure. To conclude these three experiments, each tool and technique have successfully resulted in evidence acquisition, but each requires substantial trust in the cloud infrastructure

at all levels. However, Figure 4-2 shows the results of the three experiments that acquire cloud-base evidence using today's popular tools.

Experiment	Tool	Evidence collected	Time (hrs)	Trust required
1	EnCase	Success	12	OS, HV, Host, Hardware, Network
1	FTK	Success	12	OS, HV, Host, Hardware, Network
1	FTK Imager (disk)	Success	12	OS, HV, Host, Hardware, Network
1	Fastdump	Success	2	OS, HV, Host, Hardware, Network
1	Memoryze	Success	2	OS, HV, Host, Hardware, Network
1	FTK Imager (memory)	Success	2	OS, HV, Host, Hardware, Network
1	Volume Block Copy	Success	14	OS (imaging machine), HV, Host, Hardware, Network
2	Agent Injection	Success	1	HV, Host, Hardware, Network
3	AWS Export	Success	120	AWS Technician, Technician's Host, Hardware and Software, AWS Hardware, AWS Software

Figure 4-2 Results of Three Experiments Acquiring Cloud-Based Forensics Evidence (Dykstra & Sherman, 2012)

Also, there are four alternative methods proposed by (Dykstra & Sherman, 2012) to acquire cloud-based data including: the cloud management plane, Forensics-as-a-Service and contract support. The following paragraph briefly discusses each solution:

A. Collection from the management plane: a good example is Amazon which provides a web-facing system which is called AWS Management Console. It supports trustworthy forensics by giving the consumers the user-driven ability to manage virtual assets, an out-of-band channel that interface with cloud infrastructure. The provider, customer and the third party can download log files, disk images and packet captures. Despite to this benefit, it still does require trust as a web-facing interface can open a new attack vector. Thus, it is essential to provide a secure channel between the management plane and any other endpoint by deploying robust security protocol such as Transport Layer Secure (TLS).
B. Forensics support as a service: Cloud Providers have the control on the infrastructure, including a virtual machine, logging mechanisms, packet captures and billing records. Thus, it is natural that they provide support for forensics acquisition. Once the search

warrant is issued, law enforcement requests specific data from the provider. Cloud providers can gather the data requested including virtual machines, access logs machines, access logs from the management console, data provenance logs, net flow records for the requested IP and firewall logs then send it back to the requestor. However, the main drawback of this solution is the response time and the lack knowledge of the providers about how the customer is using the cloud.

C. Legal solution: supporting law enforcements by making the options for forensics collection available is crucial in order to solve the crime. However, CSPs do not publicly announce that a forensics collection is available to law enforcement. For example, in USA context, it is not known if the Communications Assistance for Law Enforcement Act (CALEA) codified a federal law that supports law enforcement in such way that applies to cloud computing. There is a range of unique legal issues raised by the application of current law to cloud computing, specifically for the seizure of data from CSPs including who is the owner of the data and what are the jurisdictions related to cloud-based data (Forenisc Focus, 2015). These issues are including which laws governs cloud data and which is legally can execute the warrant. Dykstra & Sherman (2012) stated that they have begun analysing these legal issues in order to empower legal practitioners to understand how cloud crimes relate to traditional computer crimes and give them the tools to prosecute such cases.

   Although, Dykstra & Sherman (2012) have proposed the steps in performing the forensics investigation in a cloud environment, their work has its limitations which have to be taken in account for the future work. For example, their work is limited to IaaS using EC2 and is not suitable for other platforms and environment.

   Thethi & Keane (2014) carried out a hypothetical case study and investigated a reduction in the time of acquisition a cloud environment. Different approaches and

methods to complete remote acquisition from virtual machines in the cloud and a statistical analysis of time were used. Furthermore, partial, completed, and hybrid acquisition were investigated. The following tools were used in this experiment:

1- FTK Remote Agent and FTK imager (Mount VM)
2- FTK Remote Agent
3- FTK Imager Lite
4- FTK Imager Lite (Transferred to VM)
5- Snapshot functionality (AWS) and dd command in Linux.

Comparing to the conventional method of imaging remote data (using FTK imager lite from an external drive), all methods of data acquisition present a reduction in time of acquisition except the first method (Mounting the VM and then imaging). Furthermore, using the powerful computational resources of the cloud leads to a significant (77%) reduction of acquisition time. However, this approach is limited only for acquiring persistent storage. Thus, FTK Imager Lite (Transferred to VM) is the most efficient approach shows a 12% reduction in acquisition time. Moreover, as the partial approach does not contain permanently deleted files or information from unallocated space, thus it reduces the total time required to acquire evidence comparing to completed approach. However, the partial approach can be questionable or inadmissible in the court of law. Therefore, a hybrid approach is the most suitable approach. Figure 4-3 shows the relationship between times taken for Image acquisition and different storage capacities of Virtual Machines in the Amazon EC2 Cloud.

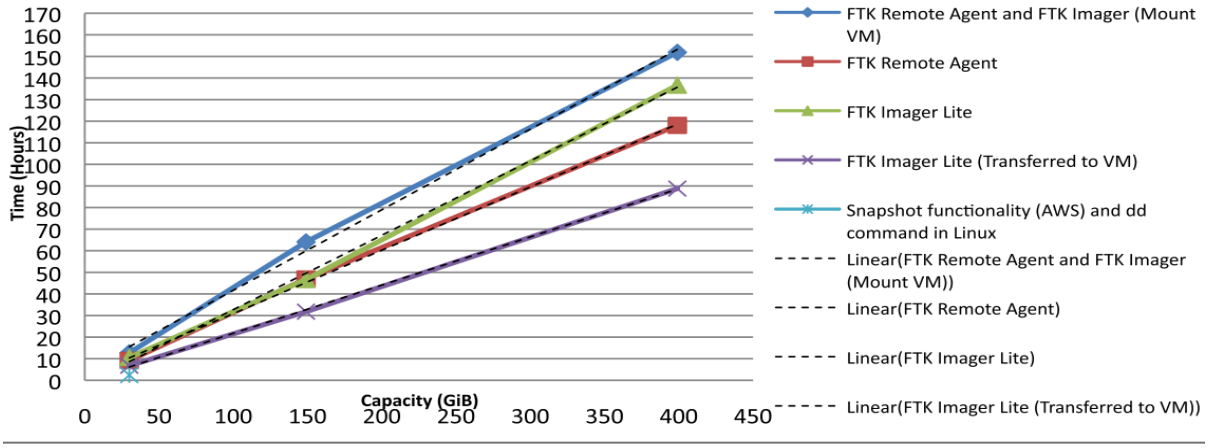


Figure 4-3 Scatter Plot Representing the Relationship between Times Taken for Image Acquisition and Different Storage Capacities of Virtual Machines in the Amazon EC2 Cloud (Thethi & Keane, 2014)

Furthermore, logs are a crucial component that the digital investigator refers to when performing digital forensics investigation. Logging is a systematic record of an object action or object status that has been taken. However, there are many researchers who proposed different models in order to log user activities that have been taken inside cloud infrastructure. For example, Sibiya et al. (2012) have proposed a standardized way to do logging from locations of interest, aiming at establishing a single and centralized logging collector and processor. It is noteworthy that intelligence should be applied on log data in order to discover relationships between data and a suspected user. Zawoad et al. (2013) have proposed their solution called SecLaaS, which provides logs for forensics purposes and at the same time it maintains the confidentiality of the cloud users. Their solution was run on OpenStack and it was found that it is feasible to integrate with the cloud computing infrastructure. In the same context, Alecsandru & Patriciu (2014) created a framework that aims to provide the digital forensics investigators with a reliable and secure method in order to monitor user activity over a cloud infrastructure; thereby implementing a logs system that runs on top of new or existing Cloud Computing infrastructures.

4.5 Challenges and Solutions along the Cloud Forensics Process

The evolution of cloud forensics is still in its infancy although cloud computing has been utilised in the market for many years (Zawoad & Hasan, 2013a). The current methodologies, procedures, tools and architectures are not designed to handle and assist the respective digital forensics in cloud environments; notwithstanding, the fact of on-going and proactive investigations are becoming mandatory components for enterprises (Poisel et al., 2012). Therefore, solutions that provide cloud forensics must be sought. To date, the researchers have focused on identification of the foremost issues that face digital forensics investigators when performing digital investigation in cloud computing. This section examines the challenges in cloud forensics that are identified in the current research literature. Furthermore, it explores the current research proposals and technical solutions addressed in the respective research. Ultimately, it highlights the open problems that need further efforts to be tackled.

Depending upon each cloud service model different issues can be encountered during a digital investigation process (Sang, 2013). Gartner, however, warned that it would be impossible to perform investigation and discovery in the cloud environment (Brodkin, 2008). Nonetheless, several conceptual solutions are proposed to overcome Gartner's assumption. In general, a digital forensics process contains four main stages: identification, preservation and collection, examine and analysis and presentation stage and the section categorises the cloud forensics issues according to these steps.

4.5.1 Identification step

An initial identification of machine(s) wherein illegal activities could be carried out and a forensics investigation is required. Due to the dynamic nature of the cloud infrastructure,

several obstacles that hinder the investigators to undertake this step exist, as discussed in the sub-sections that follow.

- ***Access to the evidence in logs***

  It is a common understanding that the identification of evidence via various sources could be challenging within the cloud environment (Birk, 2011; Dykstra & Sherman, 2011; Wolthusen, 2009). Indeed, for certain cases, investigators do not even know the location of the data due to the distributed nature of cloud (i.e. data is distributed among many hosts in multiple data centres) (Reilly et al., 2011). The availability of system status and logs files is depending on the cloud service model. It is not feasible in SaaS and PaaS models due to the limited access which the client has; whereas it is partly applicable in the IaaS model as the client has access to the Virtual Machine (VM) which behaves like an actual machine (Birk & Wegener, 2011). A number of tools and procedures which can be utilized to identify and then acquire digital evidence from the cloud have been proposed and developed. Nonetheless, the majority of them have focused merely on accessing to evidence in logs in order to trace details of the past events (Zawoad & Hasan, 2013b). Zaferullah et al. proposed and developed a standard logging mechanism that ensures the generation and retention of logs along with a log management system which collects and correlates logs (Zaferullah et al., 2011). Their approach was evaluated within a Eucalyptus cloud environment. Monitoring and analysing tools (e.g. Snort, Syslog and Log Analyser) were used in order to monitor the Eucalyptus's behaviour and log all internal and external interactions of Eucalyptus components. From the log information, it is possible to identify crucial information such as the IP address of the attacking machine, browser type, information on the number of HTTP requests and content requested. Beside these, the number of VMs

controlled by a single Eucalyptus user can also be identified. Their experimental result shows that cloud forensics will get a better advancement if the cloud service providers (CSPs) could provide a better logging mechanism.

Sang (2013) has also proposed a log-based model which suits only the SaaS and PaaS models. This solution aims to keep a separate log on the consumer side locally and synchronous it with the CSP logs using information such as unique IDs and timestamps. Hence, it enables investigators to check user activities on SaaS without the CSP's support. However, the log content is decided by the CSPs to ensure comparability. Furthermore, in order to guarantee the authenticity of log data, an incremental hash code is used to improve the efficiency and to reduce the time of verification. In PaaS, a customised log module can be supplied to the third-party, for both the consumer and the cloud provider.

Damshenas et al. suggested that it is important to identify potential evidence from the client side only. Thus, designing and configuring build-in application logs are required in order to log potential evidence such as user communication logs. In SaaS, it can be helpful to implement the feature to check the basic logs and the status of the client's usage (Damshenas et al., 2012). However, they did not provide any detail on how this application should be implemented.

Marty (2011) devised a framework for recovering logging information during an investigation in a standardize manner: when, where and what to log. After enabling logging on all infrastructure components to collect logs, a synchronized, reliable, bandwidth efficient, and encrypted transport layer is established to transfer log from the source to a central log collector. According to this proposal, only a minimum number of fields are required to be presented in every log, including the time-stamps record, application and users, session ID, severity, reason and categorization. This proactive

approach provides assurance to forensics investigators that the data is reliably generated and collected. However, this framework does not deal with volatile data which may contain potential evidence.

An encrypted logging model that logs data and then sends them to a central logging server under the control of the customer was proposed by Birk & Wegener (2011). They suggested that a mechanism that prevents potential eavesdroppers from viewing and changing the content of log during the transmission process is required. They also proposed that the CSP can provide network, process, and access logs through a read-only API to get necessary logs from all the three cloud service models.

- ***Volatile data***

When the power is turned off, volatile data cannot sustain. Likewise, when a VM is turned off or restarted, all the data will be lost unless the image is stored somewhere. Unfortunately, the existing structure of CSPs does not provide a persistent storage to the customer. Although IaaS has some advantages over SaaS and PaaS, volatile storage can be a problem unless the data is synchronised in persistent storage. Thus, volatile data that resides within the virtual environment including registry entries and temporary internet files are likely to be lost when the IaaS's customer restarts their machines (Zawoad & Hasan, 2013a; Taylor et al., 2010; Guo et al, 2012; Reilly et al., 2011). If the inspected cloud-hosted virtual machines do not have persistent storage the only option to conduct inspection and analysis is the live forensics approach (Martini & Choo, 2012). Damshenas et al. proposed the solution that provides persistent storage for the client's data. This extra storage can be utilized in data-recovery, data-safety for client and ease the data collection for investigators. For this reason, it should be globalized between CSPs in order to provide the Clients with their persistent storage.

However, it is not common for small and medium scale business organizations to employ this option due to the cost issue (Damshenas et al., 2012). Furthermore, Birk & Wegener proposed a solution to overcome the problem which is posed by volatile data (Birk & Wegener, 2011). They suggested a continuous data synchronization of the volatile data between the VM and the persistent storage. However, this approach did not provide any guidelines or practical implementation of the procedures. (Zawoad & Hasan, 2012).

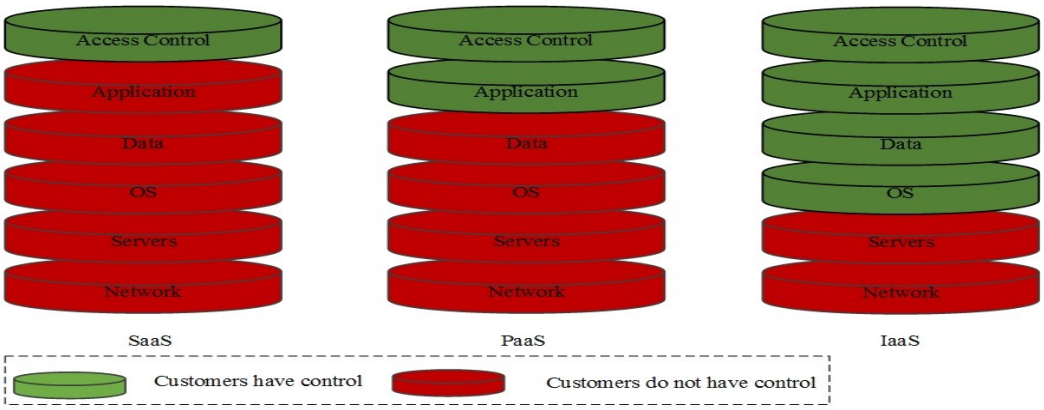


Figure 4-4 Customer Control with Different Service Models (Zawoad & Hasan, 2013b)

- ***Lack of control of the system***

The lack of control of the system poses a number of obstacles to digital investigators when they carry out the evidence acquisition (Zawoad & Hasan, 2012). Indeed, consumers have limited access and control at all levels within the cloud environment (as shown in Figure 8) and have no knowledge where its data is physically located (Dykstra & Sherman, 2013). This effectively removes the opportunity to perform a physical acquisition of the disk, which is a standard practice in computer forensics investigations. Moreover, the investigator has to obtain vital information from abstracted resources in order to accurately understand the environment including the cloud architecture, hardware, hypervisor and file system. Unfortunately, in today's

cloud architecture such information is not available to the cloud consumer yet (Dykstra & Sherman, 2011).

- ***Lack of customer awareness***

  Finally, a lack of CSP transparency along with little international regulation leads to loss of important terms regarding forensics investigations in the Service-level Agreement (SLA). This issue is applicable to all three services models (Ruan et al., 2011).

4.5.2 Data Collection and Preservation Step

This step is to collect artifacts of digital evidence and supporting material that is considered of potential value. It ensures that original artifacts are preserved in a way that is reliable, complete, accurate, and verified (Sibiya et al., 2012). However, several issues exist when investigators conduct this step in cloud-based investigations and they are listed below

- ***Dependence on cloud forensics providers***

  Both customers and investigators are heavily dependent upon the CSP in collecting the digital evidence from the cloud computing environment as they have limited control on the system. This dependence introduces serious issues of the CSP's trust and evidence integrity. Furthermore, technically, there are many reasons that prevent a CSP from providing the consumer with the desired evidence in a forensically sound manner and a timely fashion. These include but are not limited to:

  i. Due to the sheer volume of data and users within the cloud environment, most CSPs will only keep a limited amount of backups. This can cause problems when recovering deleted data or even overwritten data that is deleted by another user.

Furthermore, due to the cost and time involved in acquiring the forensic image, some cloud providers will not provide evidence beyond 1TB despite a court order served on them (Edwards, 2015). Assuming they would be willing or are required to by law, the evidence collected is still questionable as there is no way to verify the validity of evidence and whether evidence has already been lost.

ii. CSPs usually hide the data location from customers for data movement and for replication reasons (Ruan et al., 2011).

iii. In the case of an incident, the cloud provider will focus on restoring the service rather than preserving the evidence and handling it in a forensically sound manner. Furthermore, some CSPs may not report the incident or cooperate in an investigation due to potential damages upon their reputation (Wilson, 2011).

iv. CSPs do not hire certified forensics investigators to handle cloud-based incidents in a forensically sound manner. Hence, the integrity of evidence could be questioned in the court of law (Taylor et al., 2011).

v. The location uncertainty of the data makes the response time to an e-discovery request extremely challenging (Almulla et al., 2016; Ruan et al., 2013).

vi. Ultimately, as evidence is residing in one CSP, this could lead to a single point of failure and adversely impact on acquisition of useful data (Crosbie, 2013). For example, in the case of provider failure there will be significant issues in terms of services and investigation as well (Rimal et al., 2009). However, there were several failover incidents occurred so far which illustrated in Table 4-1.

Service and Outage	Reason	Duration	Date
Amazon, Netflix, Twitter, The Guardian and CNN	DoS Attack	Most of the Day	Oct 21,2016
Gmail: unavailable site because	The outage in contracts system	1.5 hours	Aug 11.2008
Google AppEngine partial outage	programming error	5 hours	June 17, 2008
S3 outage	Authentication service overload leading to unavailability	2 hours	Feb 15, 2008
S3 outage	Single bit error leading to gossip protocol blow-up	8 hours	July 20, 2008
FlexiScale	Core network failure	18 hours	Oct 31, 2008
Microsoft Azures malfunction in Windows Azure	Unknown	22 hours	March 13-14, 2008
Gmail and Google Apps Engine	Unknown	2.5 hours	Feb 24,2009

Table 4-1 Outages in Different Cloud Services

Fundamentally, the CSP architecture is designed for operational considerations to provide the most effective use of resources in the most economical fashion. As a result, they are not designed with forensics acquisition and analysis in mind. Currently, cloud customers and investigators have to completely rely on the CSPs to provide digital evidence through centralized administration and management (Ruan et al., 2013). The lack of transparency between the CSPs and customers might affect their trust relationship. Ko et al. (2011) proposed a detective model called TrustCloud which consists of five layers of accountability including system, data, workflow, policies and regulations Furthermore, Dykstra & Sherman (2012) proposed a six-layer model for IaaS based upon the amount of trust required: Guest application, Guest OS, Virtualization, Host OS, Physical hardware and Network cloud layer. The further down the stack is, the less cumulative trust is required. For example, the Guest application requires trust from all aforementioned layers, whereas the network layer only needs trust in the network. Ultimately, they recommended a cloud management plane for

using in the IaaS model in the way that customer and investigators can collect vital digital evidence including VM image and logs of network, process, and database. However, this approach needs an extra level of trust in the management plane.

Isolating a Cloud Instance

It is generally accepted in the digital investigation community that Isolating is an integral part of a Forensic process (Delport et al., 2011). For any forensics process, it is vital to isolate the incident environment in order to prevent any possible evidence tampering, alteration or damaging. Hence, it is also needed to isolate a particular instance that is connected with the incident in the cloud environment. However, achieving such a task in the cloud environment is not a trivial task due to that data instance sharing storage with multiple instances. Furthermore, a single cloud node can contain several instances and the nodes have to be cleared when performing digital investigation as a forensic investigation should not impact the availability and privacy of these users' resources (Chow et al., 2009). Users of cloud instance expect that there their privacy is protected (Lu, Lin, Liang, & Shen, 2010). In order to protect the data on that instance, some cloud isolation techniques were proposed by Delport et al. (2011) that can be used to isolate these cloud instances and mitigate the issue of multi-tenancy in cloud computing. The goal is to prevent any contamination or tampering of the evidence while forensic investigations are undertaken in the cloud environment. These techniques are Instance Relocation, where an incident can be moved into the cloud. The movement can be manually carried out by the cloud administrator or can be performed automatically via the operating system. To move the instance, it needs to be divided into three units include data on secondary storage, the content of the virtual memory and the running processes. The second proposed technique is Server Farming, which

can be used to re-route the request between user and node. The last technique is to place isolating evidence in a Sandbox. In order to obtain a better result, a combination of these techniques should be implemented. However, these techniques are mainly theory based without the support of practical experimentation.

- ***Data Provenance in Cloud***

  Provenance plays a major role to the success of data forensics in cloud computing. Implementing secure provenance enables the digital investigators to obtain vital forensics data from the cloud environment, such as defining who owns the data at a given time, and when and by whom the data was accessed. Furthermore, it maintains the chain of custody as it provides the time line of evidence. Li et al. (2013) proposed the need for a secure provenance in cloud computing that records ownership and process history of data objects in cloud computing. They stated that such techniques should satisfy conditional privacy preservation. The technique also provides confidentiality on sensitive documents stored in a cloud, anonymous authentication to cloud servers, and provenance tracking on disputed documents.

- ***Data integrity***

  One of the main issues faced by investigators in cloud based cases is data preservation (Taylor et al., 2011). Data integrity is a critical component of the forensic process (Almulla et al.,2013; Dykstra & Sherma, 2013). It is crucial that the original evidence is not changed at all (Reilly et al., 2011). A piece of incident related information has to be listed in the chain of custody register in order to maintain the integrity of the digital evidence, including how, where and by whom the evidence was collected, how the evidence was stored and preserved along with any related details of carried out

procedures (Damshenas et al., 2012). An improper preservation of evidence might become valueless in the court of law (Fei, 2007). However, it is likely that errors would occur in the data preservation stage in the cloud context due to multiple actors who are involved in the process (Zawoad & Hasan, 2013b). Thus, it is a challenging task to prove the integrity of cloud-based evidence to the court in an admissible manner (Zawoad et al., 2013). For example, if the client was involved with the malicious activities, she can claim that her authentication credentials were stolen and might be misused by somebody else. Yet, it is difficult to evaluate the authenticity of that claim (Biggs & Vidalis, 2009).

With the aim to preserve the integrity and confidentiality of the data within the cloud environment, a Trust Platform Module (TPM) was proposed by (Birk & Wegener, 2011; Dykstra & Sherman, 2012). Using TPM leads to preserving the integrity and confidentiality of the data. Furthermore, utilizing TPM solutions provides machine authentication, hardware encryption and signing, secure key storage and attestation. Beside this, it can provide the integrity of running virtual instance, trusted log files and trusted deletion of data to the customer (Zawoad & Hasan, 2013a). However, the security of the TPM is still questionable due to the possibility of modifying a running process without being detected by the TPM (Dykstra & Sherman, 2012). In the near future, CSPs are unlikely to comply with the TPM as most of the current devices are not compatible (Zawoad & Hasan, 2013a).

Furthermore, in order to authorize the client and ensure the confidentiality and integrity of the evidence, multi-factor authentication methods and cryptographic tunneling protocols such as Virtual Private Network (VPN) can be used together with an (SLA) contract should contain all clients' privacy data. Yan (2011) proposed a framework that images the relative records and files completely. Furthermore, a litigation hold or

similar freezing mechanism is required to be placed by the CSP on the account and prevent any changes to the data (Martini & Choo, 2012). For example, law enforcement agencies in Australia can make preservation notices to the CSPs according to the Australian Cybercrime Legislation Amendment Bill 2011 (Catryna, 2011). As the security is a major concern in cloud environment, researchers have proposed an encryption mechanism to ensure end user security. While this can increase the complexity of the investigation, it can also be advantageous for investigators. For example, the deployment of the Public Key Infrastructure (PKI) would be used to track down a particular suspect.

- ***Time Synchronisation***

  The synchronization of timestamps is very important as it can be used as a source of evidence. Nevertheless, the date and time stamps of the data are questionable when they are from multiple systems (Zawoad & Hasan, 2013a). Moreover, the difference in time zones between cloud servers and cloud clients can affect the integrity, reliability, and admissibility of evidence. Currently, the cloud infrastructure is strongly dependent on whether the VM guest OS is using a network protocol to synchronize with a network time server. However, the best strategy recommended by (Marangos et al., 2014) is to obtain the time from many servers and keep the most common time value from them. Furthermore, using a specific time system such as GMT on all entities of the cloud can be helpful in providing a logical time pattern in the way that enable investigators to create the time-line analysis and to track multiples log records in different physical locations (Damshenas et al., 2012).

- ***Cloud literacy of investigators***

  There is a lack of training material that educates investigators on cloud computing technology and cloud forensics procedures. Current digital forensics training materials are not updated nor do they address the major challenges of cloud environments. Beside this, there is lack of standard operating policies for cloud forensics (Ruan et al., 2011). It is highly needed that members of an investigation team should be trained legal law regulations, special tools and techniques, including programing, networking, communication and negotiation with CSPs (Chen et al., 2012).

- ***Chain of custody***

  The chain of custody is one of the most critical problems in the digital forensics arena (Zawoad et al., 2013). The chain of custody has to illustrate how the evidence was collected, analysed and preserved with the aim of presenting the evidence in an admissible way at the court of law (Dykstra & Sherman, 2011). It is difficult to verify the data chain of custody in the cloud environment, due to the unique combinations of characteristics that the cloud computing has, including the distributed and multi-layered nature (Ruan, 2013). In order to maintain the chain of custody, certain things are required to be clarified, such as the way in which logs were collected, generated and stores along with who had the access to the logs. Moreover, CSPs have to hire trained and qualified specialists (Grispos, 2012). Furthermore, communication and collaborations related to all forensics activities through the chain of CSPs and customer's dependencies need to be clearly written in SLAs (Ruan et al., 2011).

4.5.3 Analysis and Examination Step

It is very challenging to conduct a proper analysis in the cloud due to the sheer volume of resources and vast objects to be examined in the digital investigation along with limitation in processing and examining tools. Moreover, there is no standard program for the forensic extraction of data as the customer can access relevant data from various devices such as desktop PC, tablet and mobile phones and a wide range of applications. Furthermore, the data extraction format is varied based on the service model. For example, in the IaaS model, investigators can obtain an image of the virtual machine that contains all data uploaded by a suspect. However, the data would be exported in an unstructured fashion, creating difficulties in reading, examining and analysing the data format by using standard forensics tools. Thus, it is important to develop utility applications that translate the native cloud data format to a readable and recognizable format (NIST, 2011). Reconstruction of events in a forensics investigation produces crucial and valuable analysis in order to logically recreate the crime. However, due to the distributed and shared nature of the cloud, each event of the crime might occur in different countries. This will lead to the difficulty in making the logical order of where it took place. Investigators could face a wide range of challenges when they perform the examination and analysis stage, including:

- ***Lack of Forensics Tool***

  There is consensus of most of the researchers that the available forensics tools cannot cope up with the distributed and elastic characteristic of cloud computing. Several researchers highlighted the limitations of the current tools in their research (Ruan et al., 2011; Reilly et al., 2011; Grispos, 2012). There is also a high level of demand for forensic-aware tools for the CSP and the clients to conduct forensics investigation in a cloud environment (Ruan et al., 2011). Hence, it is crucial to develop tools which can

be utilized to identify, collect and analyse forensics data in cloud environment (Shah & Malik, 2013).

A combination of computer forensics and network forensics tools is needed in order to acquire forensics data and then analysing them in a timely fashion. Traditional forensics tools can be used to collect the active data while its integrity is preserved. Network forensics tools can be utilized to collect additional data over the network including activity logs (Zargari & Benford, 2012). E-discovery refers to any process in which electronic data are sought, located, secured for the purpose of using it later in the legal case. In the cloud computing environment, E-discovery can be helpful to conduct offline investigations on a particular computer or network. For example, Encase software has launched their own e-discovery suite; nevertheless, a multi-jurisdiction problem is still a major concern (Biggs & Vidalis, 2009). In cloud computing, it is less likely that CSPs will obey the legal e-discovery obligations due to technical, cost and legal reasons or even to incapability to preserve the original metadata as expected (NIST, 2011). Furthermore, the response time to an e-discovery request is extremely challenging due to the uncertainty of data location and the need for assurance of completion of the request (Ruan et al., 2013).

The open-source software, Offline Windows Analysis and Data Extraction (OWADE) were developed and launched at the BlackHat 2011 conference by researchers from Stanford University in California. This software has the ability to find out which website a user visited, extract information stored in cloud, reconstruct Internet activities and search for the online identities that were used. This version is still under development and it only works against Windows XP drives (Kumar, 2011). Furthermore, the management plane was recommended as the appropriate forensics tools for acquiring cloud-based data (Dykstra & Sherman, 2012). They claimed that

management plane offers the most attractive balance between speed and trust. Despite the fact that some commercial tools (e.g. Encase and FTK) can be used to successfully acquire evidence, Dykstar et al. do not recommend them due to the high level of trust they required. Recently, Dykstra et al. developed a management plane forensics toolkit called Forensics Open-Stack Tools (FROST) which is designed to acquire forensics data from virtual disks, API logs and guest firewall logs (Dykstra & Sherman, 2013). It operates at the cloud management plane instead of interacting with the operating system inside the guest virtual machines. FROST is the first forensic tool that is built into any IaaS cloud model (Dykstra & Sherman, 2013). However, FROST is deployed by the CSP. Thus, trust in the CSP is required but not in the guest machine. Furthermore, trust on the cloud infrastructure is required including the hardware, host operating system, hypervisor and cloud employees. FROST also assumes that cloud customer is cooperative and involved in the investigation.

- ***Evidence Correlation across Multiple Sources***

  Correlation of activities from multiple sources can be overwhelming. The resources of evidence are spread across multiple digital resources. Handling data evidence from multiple sources introduces a problem for investigators.

- ***Crime Scene Reconstruction***

  It is crucial to reconstruct the crime scene in order to understand how illegal activities were committed. Unfortunately, this could be a problem in the cloud environment (Zawoad & Hasan, 2013a). For example, when an adversary shuts down their virtual instance after committing certain malicious activities, reconstruction of the crime scene will be impossible. However, regeneration of events can be used where a snapshot is

performed due to the occurrence of every attack. Geethakumari & Belorkar (2012) proposed a method allowing investigators to replay the event of the attack and restore the system to the state before the attack by using snapshots. Ultimately, it is also suggested that incoming and outgoing data through the cloud is visualised by the investigators. Almulla et al., (2013a) demonstrated that it is possible to extract evidence from distributed snapshot dumps in a forensically sound manner. In 2016, they developed a forensic process model that successfully examine snapshots and acquire forensic data without the need to reconstruct the virtual environment (Almulla et al., 2016). This work was conducted at a small scale which did not provide a comprehensive insight of the proposed system thus they intended to perform a further study on the big data distributed storage snapshot.

4.5.4 Presentation Step

The final step of digital forensics investigation is presentation, where the evidence has to be presented to a judicial body in the form of a report or testimony (Trenwith & Venter, 2013). Several challenges lie in this step in context of cloud forensics. For instance, it is not clear how to specify the physical location of the cloud-based crime due to distributed and shared resources between multiple clients who are based in different countries. This, in turn, confuses the investigators to determine which legal system the case should be heard. Furthermore, it is required that digital investigators have to technically explain to the jury how the evidence was acquired and what it represents. However, the technicalities of a cloud data centre, running thousands of VMs, accessed simultaneously by hundreds of users are very hard to be comprehended by a jury member who is likely to have a basic technical knowledge (Reilly et al., 2011).

4.6 Discussion of the Cloud Forensic Solutions

It is clear that there are plenty of issues that need to be tackled in order to perform a proper forensics investigation in the cloud environment. Table 4-2 illustrates challenges and their potential solutions. All proposed solutions were identified from the review conduced in the respective domain. Table 4-3 summarises the open problems that need to be resolved. Several solutions were proposed with the aim of mitigating cloud challenges. Despite this, the majority of these solutions are conceptual and not tested in real conditions. So far, traditional tools such as Encase and FTK are still the common tools that are heavily utilised in acquiring the evidence from the cloud despite the difference between conducting digital investigation in the cloud infrastructure and in tradition computer environments. According to the previous section, there was only one research that evaluated and examined the current tools used in conducting remotely data acquitting. This research conducted by Dykstra and Sherman who developed a set of tools known as Forensics OpenStack tools (FROST). It operates at the cloud management plane instead of interacting with the operating system inside the guest virtual machines. FROST is the first forensics capability to be built into any Infrastructure-as-a-service cloud model. However, FROST is deployed by the CSP. Thus, trust in the CSP is still required but not in the guest machine. Furthermore, trust on the cloud infrastructure is required including the hardware, host operating system, hypervisor and cloud employees. It also assumes that the cloud customer is cooperative and involved in the investigation. This work has performed three experiments to acquire forensics data from three different layers namely guest OS, the virtualisation layer and the host OS. All three experiments have succeeded in performing data acquisition remotely from the cloud-based layer. However, a certain amount of trust is highly required in each layer. Data acquisition is the first practical required task that digital investigators have to begin with. Customers and investigators depend on the CSP to conduct this task. Few researchers suggested solutions that would mitigate the issue of the dependence on the CSP such as cloud

management plane or API which are provided to the customer in order to get forensics data. However, there is various and crucial forensic data that still reside in the CSP including deleted files from the hard disk and temporary registry logs. In order to acquire this kind of data, relying on the CSP cooperation is currently inevitable. In turn, many others issues associated with the dependence on the CSP evidence are shown and they are not resolved yet. Such issues are including trust, delayed response, inadmissibility of evidence and potential single point of failure. Furthermore, piecing together a sequence of events from multiple sources and different jurisdictions is another major obstacle faced by investigators in the cloud environment. So far, investigators have no valid approach to reconstruct the past state of event with a level of accuracy that the reconstructed information can be admissible in the court of law. Several difficulties associated with logging data are still not diminished yet. These are including time-line, log review, logging correlation and log policy monitoring. Ultimately, legal issues hinder the smooth performing of forensics investigation due to the lack of guidelines and implementation of global unity to overcome the cross border issue.

Cloud Forensics challenges/ Process		Apply to Service model			Potential Solution	Ref
		IaaS	PaaS	SaaS		
Identification						
Access to the evidence		√	X	X	Eucalyptus framework OS and the security log	Zaferullah et al. (2013)
		√	X	X	a log-based model	Sang (2013)
		√	√	X	Extraction of relevant status data	Damshenans et al (2012)
		X	√	X	A log management solution	Mart (2011)
		√	√	X	An encrypted logging model	Birk et al (2011)
Dependence on CSP for	Trust issue	√	√	X	Layers of Trust Model	Dykstar et al (2012)
		√	√	X	TrustCloud	Ko et al. (2011)
	Data acquisistion	√	√	√	Cloud Management Plane	Dykstra (2012)
	Compliance	√	√	√	SAL	Dykstra (2011), Biggs & Vidalis (2009)
	Logs	√	√	√	API provided by CSPs	Birk et al (2011)
Lack of customer awareness		√	√	√	Dedicated awareness programmes	Ruan et al. (2011)
Volatile data		√	√	X	Client Persistent Storage	Damshenas et al. (2012)
		√	√	X	A continuous synchronisation API	Birk et al. (2011)
Lack of system control		X	X	X	--	--
Preservation & Collection						
Data integrity		√	√	√	Trust Platform Module (TPM)	Birk et al (2011), Dykstra (2012)
		√	X	X	A Distributed Snapshot Framework	(Almulla et al., 2016)
Time synchronisation		√	√	√	Unified/specific time system	Damshenas et al. (2012)
Cloud literacy of investigators		√	√	√	Developing investigators technical skills	Chen et al., (2012)
Chain of custody		√	√	√	Trained staff	Grispos (2012), Ruan et al. (2011)
Analysis & Examination						
Lack of forensics tool		√	√	X	FROST, OWADE	Dykstra & Sherman (2013), (Kumar, 2011).
Evidence correlation		X	X	X	--	--
Presentation						
Jury's technical comprehension		X	X	X	--	Reilly et al (2011)

Table 4-2 Cloud Forensics Solutions

Open Issues	
1	Tackle the dependence on the cloud services providers particularly in data acquisition and trust issues
2	Timeline analysis across multiple sources and evidence correlation
3	Overcome the cross-border issues
4	Lack control of the system
5	Jury's technical comprehension

Table 4-3 Unresolved Issues of Cloud Forensics

4.7 Conclusion

As there is increasing use of cloud computing, there is a growing need for trustworthy cloud forensics. Several researchers have identified and explored the challenges confronting the digital investigators when they conduct a forensic investigation in cloud-based cases. Accordingly, few researchers have proposed technical solutions to mitigate these challenges. As such, there are still open issues need to be tackled. Dependence on the CSP is a major challenge such as trust issue, delay response, inadmissibility of evidence and potential single point of failure. Furthermore, timeline analysis across multiple sources hinders investigators to understand the relationship and data flow between systems. Unfortunately, to date, most studies are focussed upon techniques and tools involving one source of digital evidence at a time for investigation.

5 Experimental Validation

This chapter briefly introduces a novel approach which involves acquiring the virtual hard disk from the cloud environment at any given time with no requirement of the CSP to be involved or any modification to the CSP's underlying infrastructure. Furthermore, this chapter performs a series of models that experimentally proven with the aim of investigating the approach's investigating the approach's admissibility, efficiency and feasibility.

5.1 Introduction

The acquisition of digital artifacts via digital forensics is an essential step in any forensics process. As outlined in the previous chapter to date, no such system exists to support the routine forensics data acquisition and analysis of cloud systems. At present, when a consumer or law enforcement agency requires evidence, support from CSPs is essential to obtain access and acquire evidence – no matter what state that information might be in. Currently, cloud data resides in a virtual instance, and relying on the CSP is unavoidable. This, in itself, raises associated issues, including trust relationship, delay response, inadmissibility of evidence and potential single point of failure.

While previous research has proposed an IaaS solution, this fundamentally relies on the collection and storage of VM images, and a key aspect of the approach is to include the CSP as a core part of the solution – largely to ensure cloud management information is also obtained. Thus, the issue of dependence on the CSP and its associated problems are not resolved yet. This chapter briefly introduces a novel approach which involves acquiring the virtual hard disk from the cloud environment at any given time with no requirement of the CSP to be involved or any modification to the CSP's underlying infrastructure. Such an approach needs to be experimentally proven in order to see its feasibility. Thus, three experiments are conducted in

order to investigate the approach's admissibility, efficiency and to provide a better understanding of both the technical implications resulting from such a system regarding the day-to-day operation of a cloud system as well as the financial costs. In order to appreciate the need for the experiments, the core novel approach is described.

5.2 A Novel Forensic Acquisition and Analysis System (FAAS) in an IaaS Model

It is imperative that organisations remain in control of their data and have the ability to undertake incident analysis/forensics examination of their systems when required. The Platform-as-a-Service and Software-as-a-Service models are built on IaaS. Thus, beginning with IaaS provides a fundamental basis from which to build future work (Almulla et al. 2014; Dykstra & Sherman, 2012). In IaaS, the consumer has complete control over a guest operating system running in a virtual machine (VM), whereas the provider retains control and responsibility for the hypervisor (HV), down to the physical hardware in the datacentre.

The author has proposed an approach in this research entitled the Cloud FAAS (a Forensic Acquisition and Analysis System). It aims to enable the IaaS cloud customer to have complete control over the forensic acquisition process, ignoring the data held by the CSP. This is made possible through the implementation of an agent-based approach that sits on each of the customers' VMs and communicates the necessary information to a central forensic storage device and retrieves the hard disk, when needed, in both a forensic sound and timely fashion.

While researchers have suggested using built-in cloud functionality that provides a snapshot of the virtual disk, this approach captures only the allocated contents rather than the complete disk. On a physical level, the virtual hard disk in the cloud is stored upon many disks that are shared between tenants and remotely distributed. Thus, complete imaging of the virtual hard

disk on which the VM resides is usually not permitted by CSPs in order to ensure compliance with relevant laws and regulations, thereby maintaining the confidentiality of other tenants.

This warrants a system which enables the investigator to acquire an image(s) of the hard disk in a forensically admissible manner. Cloud FAAS aims to achieve this through an agent-based approach that resides on the operating VM and therefore has complete logical access to the disk (which is completely independent of where it physically resides). In doing so, it has the capability to image and monitor for changes to the virtual disk. Operating at this logical layer on a VM means it is not possible to map this data to a physical drive, indeed; the nature of the cloud infrastructure would not permit this. The agent is set to initially create a base image of the non-volatile memory (the VM OS only sees it as a cluster-based disk) and save it in the image repository. To ensure the forensic value of such data, the monitoring is performed at a cluster level, maintaining a complete record of all clusters over time.

The focus upon the cluster rather than on the files provides for a robust approach that also captures the unallocated areas of the disk. In this manner, all file changes are recorded from initial imaging and throughout the use of the VM. The process begins with a complete forensic image of the disk at start-up by the non-volatile agent, communicating the image back to the Cloud FAAS storage. The agent then proceeds to monitor the clusters and communicates all modified clusters back to the Cloud FAAS. Indeed, operating similar to an incremental backup will reduce the volume of data that need to be communicated and stored at a minimum. The complete image and cluster contents are immediately hashed and secured to ensure integrity and confidentiality of the contents. To ensure the forensics value of such data, the data clusters of those file changes are also stored. This allows an image reconstruction engine to reproduce a forensic image of the drive at any required point and provide access to every file in its entirety including showing how files have been deleted and overwritten. Doing so removes the

limitations of today's file carving tools including fragmentation and validation on the files that they carved.

5.3 Scientific Method

The proposed approach raises a number of research questions that need addressing to determine its viability. Three experiments were conducted with the aim of:

1. Investigating the possibility of monitoring the virtual hard disk at a cluster level and the ability to reconstruct the hard disk at any given time, reproducing the gold-standard forensic hashing as if it were a standard HDD acquisition. Experiment 1 seeks to monitor all data changes at a cluster level and to validate these to prove that it is possible to produce reliable forensic images (in terms of hashing) that can be reconstructed at any point in time.
2. Investigating what Cloud FAAS would look like operationally, with a certain desktop user's usages, utilising an IaaS workstation-based environment. Are the operation and cost too expensive for the approach to be feasible? Experiment 2 will explore and answer this question.
3. Looking at the nature of server-based architecture, which is typically more common with cloud-based infrastructure, including web services, application servers, database servers, active directories and file servers.

The following section describes each experiment and shows the results, followed by an overall discussion.

5.3.1 Experiment 1

This experiment seeks to validate that the Cloud FAAS is capable of producing reliable forensic images (in terms of hashing) that can be reconstructed at any point in time.

The agent is set to initially create a base image of the non-volatile memory (i.e. the hard drive as seen by the VM) and save it in the image repository. For every single cluster of the virtual hard disk a hash value is generated and saved. This assists in ensuring and maintaining integrity, but also in order to detect any data changes that have occurred on the hard disk at a cluster level. The agent also utilises metadata files to monitor any change on the hard disk (e.g. $logfile in NTFS). Once the change is detected, the agent will generate a hash for each cluster and compare them with the previously taken hash to identify/confirm that a change has taken place.

Once a change is detected, the agent seeks to operate a procedure which is similar to an incremental backup, but at a cluster level, imaging the cluster and sending both the hash and the cluster image to the Cloud FAAS repository. In order to generate cryptographic hash values, the author uses a MD5 – as it is the fastest algorithm for experimental purpose.

When image reconstruction is required, an investigator from the organisation will select the date/time stamp and the reconstruction engine will create the forensic image, as if it were forensically acquired as per normal (i.e. a complete bit-for-bit copy). Utilising the initial base image, the reconstruction engine will sequentially apply the cluster changes up to the date/time selected. A hash is then taken of the image and stored.

The experimental testbed was set up for this scenario, consisting of a VM running Windows 7 Professional on VMware Workstation 12 with 9 GB RAM, 2x CPU @ 2.5 GHz, 1 Gbps NIC

and 30 GB HDD. A non-volatile agent (developed for FAAS) is installed on this VM and linked to the Cloud FAAS architecture for storage.

Analysis of Experiment1

For experimental purposes, in order to validate if the reconstruction engine is able to reproduce a drive image as if it were a standard acquisition, a full image of the system was taken twice a day over a 10-day period. This created 20 restore points. In order to ensure sufficient disk/system activity, an IaaS system was set up and used by the author as his core work-based PC during this 10-day period – this reflected the use that would have been seen by cloud users of IaaS systems as a workstation (VMware, 2015).

In addition to validating the forensic integrity of the process, it is also important to investigate the processing, memory and storage overheads of the approach, to determine its viability. The performance overhead is analysed using Windows Task Manager, which provides key data about how a running agent uses system resources, including random access memory (RAM) and the central processing unit (CPU). In this scenario, network resource utilisation and performance are not considered; however, as the volume of data is monitored, it is possible to appreciate the likely impact upon network resources.

As illustrated in Table 5-1, the generated hash values from the hard disk immediately after the data changes match the generated hash value of the reconstructed images in all 20 samples. It is, therefore, evident that the Cloud FAAS is able to successfully reconstruct the forensic image of the disk at any specified time, ensuring the same data integrity as digital investigators already experience in computer-based forensics. Hence, such evidence is forensically sound and, ostensibly, admissible in a court of law.

#	Day	Reconstructed Image	Forensic Image	Matched
1	1	93c8b7fcf73bac84ce19418c84dda3d5	93c8b7fcf73bac84ce19418c84dda3d5	√
2	1	8391bc68c77c6c8f6ca2bbb97cb3f1cd	8391bc68c77c6c8f6ca2bbb97cb3f1cd	√
3	2	db6c16d889585cd44316761eec6a8ad0	db6c16d889585cd44316761eec6a8ad0	√
4	2	698152673c066d3f47ec1e092b77a12b	698152673c066d3f47ec1e092b77a12b	√
5	3	221a315871d068532bd2f65685ac3d05	221a315871d068532bd2f65685ac3d05	√
5	3	4c4855f04abcf93d2d394d0eb0249871	4c4855f04abcf93d2d394d0eb0249871	√
7	4	17932ae42683b1a69d76190b5a9c29a2	17932ae42683b1a69d76190b5a9c29a2	√
8	4	2d8813de6cb03717a1f9151015e907b6	2d8813de6cb03717a1f9151015e907b6	√
9	5	3b19f0e97d8fe5bee55f3bf3335537f0	3b19f0e97d8fe5bee55f3bf3335537f0	√
10	5	aae3c4a85bd8129287de72d711d6958f	aae3c4a85bd8129287de72d711d6958f	√
11	6	85c70cfd90892f0b27ccce8b598daef3	85c70cfd90892f0b27ccce8b598daef3	√
12	6	9f60ecfbe5a97ebb8a6f5a1a6e2381ae	9f60ecfbe5a97ebb8a6f5a1a6e2381ae	√
14	7	a8ce08bd6e1453e9f566355cd5f19e44	a8ce08bd6e1453e9f566355cd5f19e44	√
15	7	afef214455eb03f2b7c33f911224df20	afef214455eb03f2b7c33f911224df20	√
16	8	62c9929f1e4ced3d2771984df84754c0	62c9929f1e4ced3d2771984df84754c0	√
17	8	a041c62788a1098ef85eed2898fadc46	a041c62788a1098ef85eed2898fadc46	√
18	9	b8a50639ab2cb0df84723daf75e6212e	b8a50639ab2cb0df84723daf75e6212e	√
19	9	efab1fe380cd67d71a802dbb21a32a59	efab1fe380cd67d71a802dbb21a32a59	√
20	10	9af19402e4b83809d9b17f2209cd8de6	9af19402e4b83809d9b17f2209cd8de6	√

Table 5-1 Reconstructed Images vs. Forensic Image

Figure 5-1 illustrates the memory consumption that takes place over a typical hour. Notably, the amount of memory utilised to perform this acquisition process and monitoring is less than 10 MB. It is also noteworthy that this has a minimal impact upon the memory consumption, even during the period where large volumes of disk activity/change are occurring.

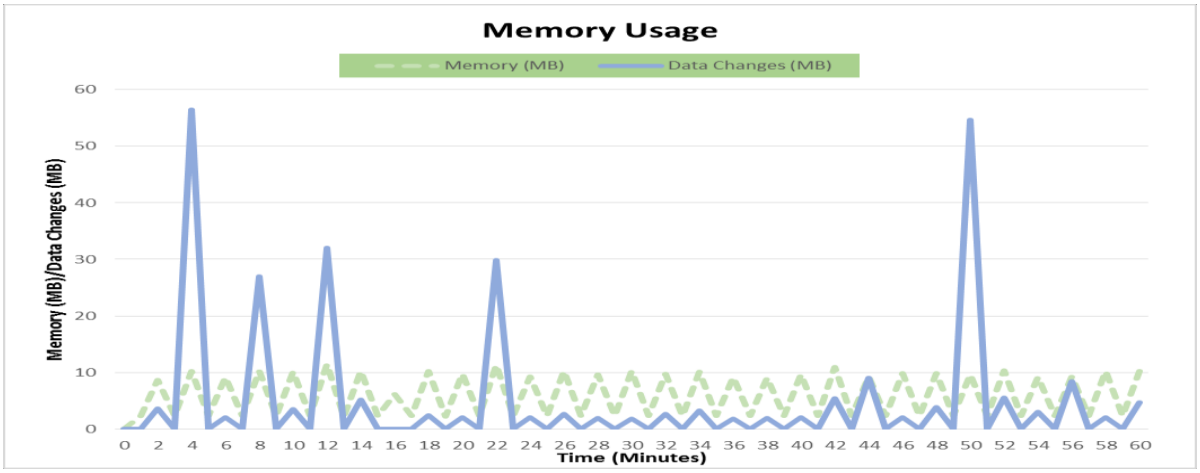


Figure 5-1 Memory Usage

Investigating the CPU utilisation across the same period shows that the agent has a larger impact (as illustrated in Figure 5-2). However, it should be noted that the acquisition agent was not designed with any functionality to specifically use unused processing cycles. If it were, it would result in a minimal impact upon the core activities of the system.

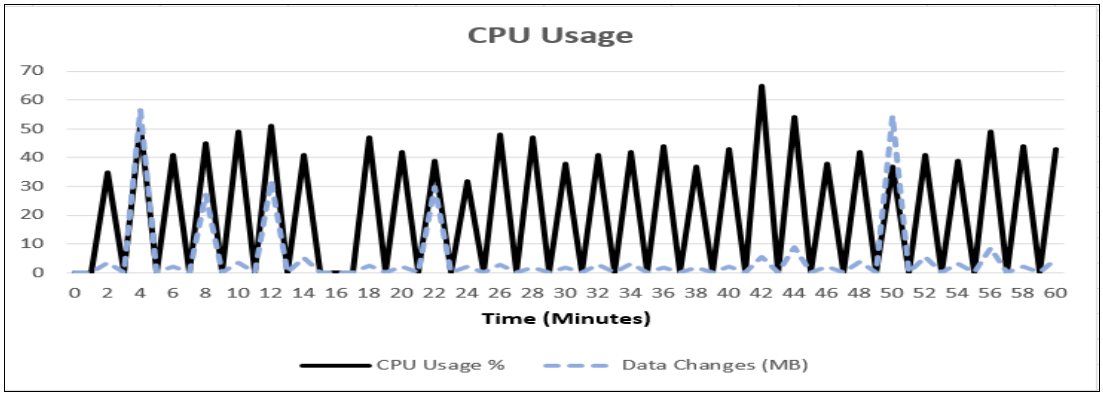


Figure 5-2 CPU Usage

The forensic image storage requirements are presented in Table 5-2. It shows the size of the disk changes in MB per day over the 10 days of this experiment. It also indicates that the total amount of storage required storing all these changes are 3655.9 Mb over the 10 days. Contrasting this acquiring approach and storing multiple complete snapshots or forensic images, this approach significantly reduces the volume of data to be stored. For example, a

30GB disk forensically imaged daily would result in 300GB of uncompressed space across 10 days. This approach requires a total of 33.65GB.

Day	Size of data changes (MB)
1	277.3
2	339.2
3	287.0
4	427.1
5	372.4
6	522.3
7	209.6
8	220.7
9	493.2
10	477.1
Total	3655.9

Table 5-2 Forensic Acquired Images Storage

In terms of network bandwidth, even scaled, a daily transfer of 200-500MB is not significant in comparison to the normal volume of daily network traffic.

5.3.2 Experiment 2

This experiment investigates what Cloud FAAS would look like operationally, in relation to a certain desktop user's usages, when utilising an IaaS workstation-based environment. It discovers whether the operation and the cost would be too expensive for the approach to be feasible.

A non-volatile agent (developed for FAAS) was installed on six VMs that deployed and assigned to six students in the university lab. All VMs are linked to the Cloud FAAS architecture for storage.

The experimental test bed was set up for this scenario, consisting of one vCloud Director Server. This server shares a common database and is linked to an arbitrary number of vCenter servers and ESXi hosts. vShield Manager servers provide network services to vCenter and

vCloud Director servers. All VMs are running Windows 8 with 2 GB RAM, 2x CPU @ 2.5 GHz, 1 Gbps NIC and 30 GB HDD. The documentation relating to the set-up of this environment is attached in Appendix I.

Analysis of Experiment 2

To gain more insight about the hourly variations in disks changes monitored by the developed agent, the distribution of data change during the day for all VMs were observed and presented in Figure 5-3.

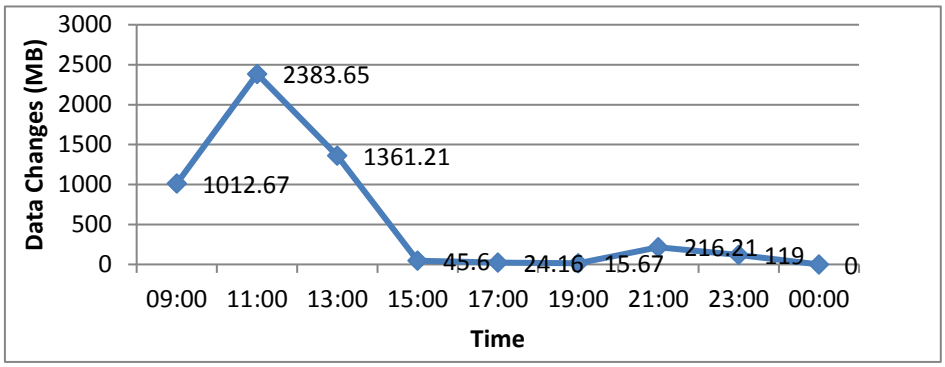


Figure 5-3 Data Changes (MB) per day for all users

According to the findings, most data change peaks twice during the day. One peak is between 10 a.m. and 1 p.m. (during working hours) and the other is in the evening, between 9 p.m. and 11 p.m.

During peak hours, the users contribute more than 92% of their activity. This is a large amount of data change that needs to be recorded at the same time, which will overload both the server as well as the users' machines.

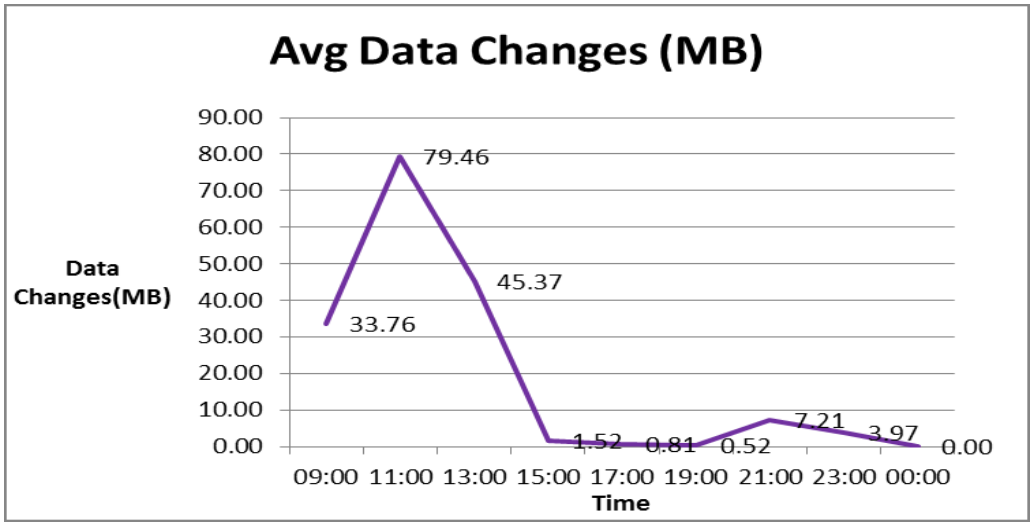


Figure 5-4 Data Changes per Day per User

To extend this calculation, the average amount of data change per day per user during different times on a typical day was also calculated. As shown in Figure 5-4, a typical user contributes 172.62 MB of data change in a typical day, with 158.59 MB during working hours (9 a.m. – 1 p.m.) and a peak of 79.46 MB at 11 a.m. These values can be multiplied by the number of users connected to the server at any given time to calculate the estimated amount of data change so that the organisation can assess the situation and plan accordingly.

Regarding the data distribution for all users daily between the hours of 9 a.m. and 11 p.m., peak data generation was less than 500 MB within a 60-minute window. This volume of network traffic is insignificant in contrast to normal network traffic. However, the impact of that volume of data traversing a network may impact the operation, depending on the network optimisation of network set-up, including traffic shaping, which delays the flow of packets that have been designated as less important or less desired than those of prioritised traffic streams, data deduplication and compression techniques.

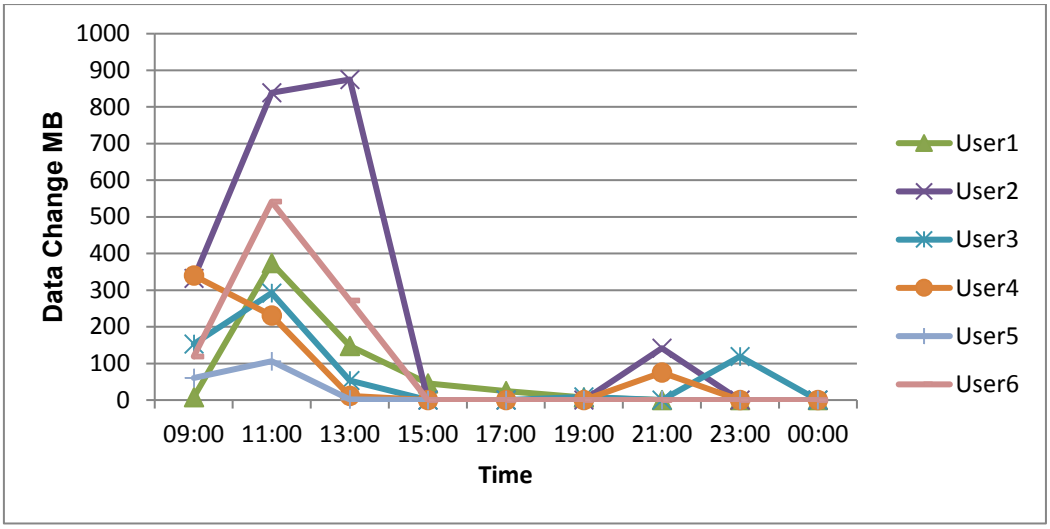


Figure 5-5 Distribution of Data Change for Each of the Six Users across All Days

Multiple time slots were observed, and activity was found to be highest between 10 a.m. and 11 a.m. as shown in Figure 5-5. Data activity, and another resource usage, was observed to be at its highest during this period; as such, the time slot was chosen to create as clear a picture as possible of the potential impact of FAAS on system resources.

In order to investigate the processing, memory overheads of the approach to determine its viability, typical VM should be selected to be investigated. As shown in Figure 5-6 below, User 4 can be regarded as the median of all the users, therefore enabling more predictable results to be gathered for the experiments discussed in this section.

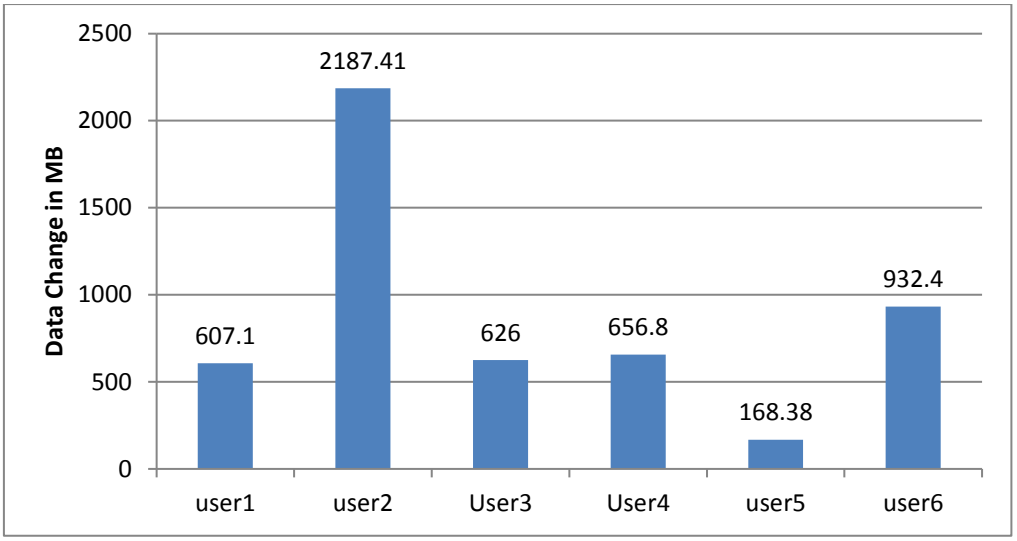


Figure 5-6 Each User's Data Activities

The performance overhead is analysed using Windows Task Manager, which provides key data about how a running agent is using system resources, including RAM and the CPU, as shown in the following analysis.

Figure 5-7 shows the CPU usage for User 4 during the peak hour of typical usage. It is obvious that the data change is sparse, with peaks in specific times during the peak hour and a maximum of approximately 80 MB data changes. CPU usage is significant (up to 50%) at all times, indicating the agent's high impact on the system.

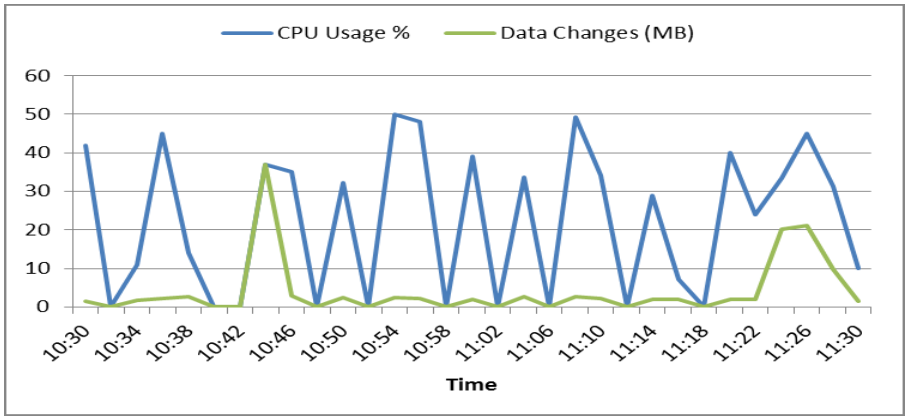


Figure 5-7 CPU Usage for the Typical User during the Peak Hour

Figure 5-7 shows an increased CPU utilisation at all times, regardless of disk changes at peak periods; it can be assumed that the agent creates an appreciable impact on system resources.

Fortunately, this is not the case for memory consumption, which is kept below 10 MB, which is low at all times regardless of the data change at peak periods, as shown in Figure 5-8. This also agrees with the results obtained from Experiment 1.

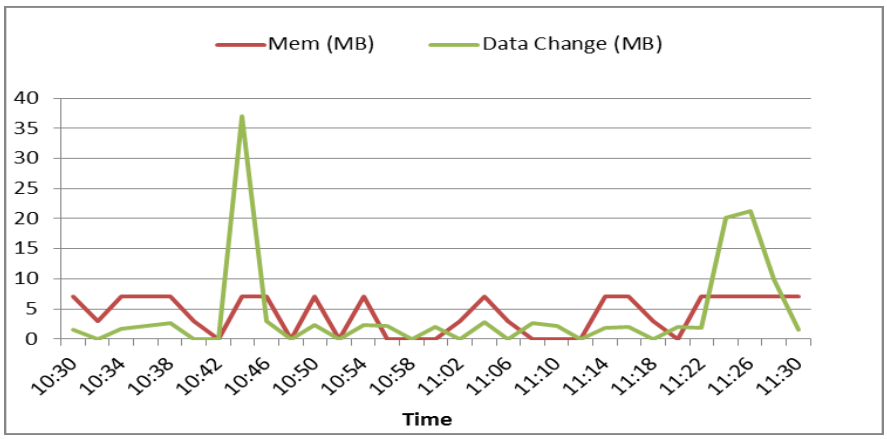


Figure 5-8 Memory Usage for User 4 during the Peak Hour

The forensic image storage requirements are presented in Table 5-3. It shows the VM role/username, the size of the VM hard disk and the average results of daily, weekly and monthly data changes. It also compares the cloud FAAS reconstruction requirements against a traditional forensic approach when it generates a complete snapshot on a daily basis.

VM role		The size of HD	Data Changes (GB)			Reconstructed Forensic Image after one month Cloud FAAS	Daily Multiple Complete Snapshots For one month
			Daily	Weekly	Monthly		
Desktop	User 1	30	0.12	0.84	3.6	33.642	900
	User 2	30	0.44	3.08	13.2	43.122	900
	User 3	30	0.13	0.91	3.9	33.756	900
	User 4	30	0.13	0.91	3.9	33.936	900
	User 5	30	0.03	0.21	0.9	30.99	900
	User 6	30	0.17	1.19	5.1	35.594	900
Average		30	0.17	1.19	5.1	35.594	900

Table 5-3 The Requirements for Forensic Image Storage (GB)

The data that has been generated by the monitored workstations, as described in the table above, shows that the size of data changes on a daily basis. Accordingly, the disks were further predicted to obtain data on a weekly and monthly basis. It is clear that the cumulative changes over the stated periods and resulting storage requirements are a drastic reduction when compared with the required storage space for a traditional full disk image (complete snapshot). To elaborate, each disk used has a storage capacity of 30 GB, and data changes were monitored across the stated time periods. The highest amount of generated data for the users was less than half a gigabyte daily (User 2 = 0.44 GB). That user's monthly data generation indicated the highest level of activity, at 13.2 GB, which is less than half the space required if traditional full disk image backups were used even for just 1 day.

Looking at a monthly basis, for instance, the average that Cloud FAAS requires to reconstruct the forensic image at any given time during a 1-month period for a specific VM is 35.594 GB, whereas the traditional approach will result in 900 GB. Thus, it can be said that the average forensic storage required to reconstruct the hard disk in a forensic manner at any given time for a month is 35.5 GB, which is only 3.89% of the size required to maintain traditional snapshot mechanisms. The data is a mere 17% (5.1 GB vs 30 GB) while still reconstructing the forensic image at any given time, maintaining data integrity as well as reducing storage costs and operational impact.

5.3.3 Experiment 3

In this section, the author aims to examine the nature of server-based architecture, which is typically more common with a cloud-based infrastructure. One of the interviewees who contributed to evaluate the Cloud FAAS was interested in sharing the real data extracted from his organisation. For marketing purposes, the name of the firm is not revealed in this thesis, but a brief description of the company's set-up is given below in Figure 5-9.

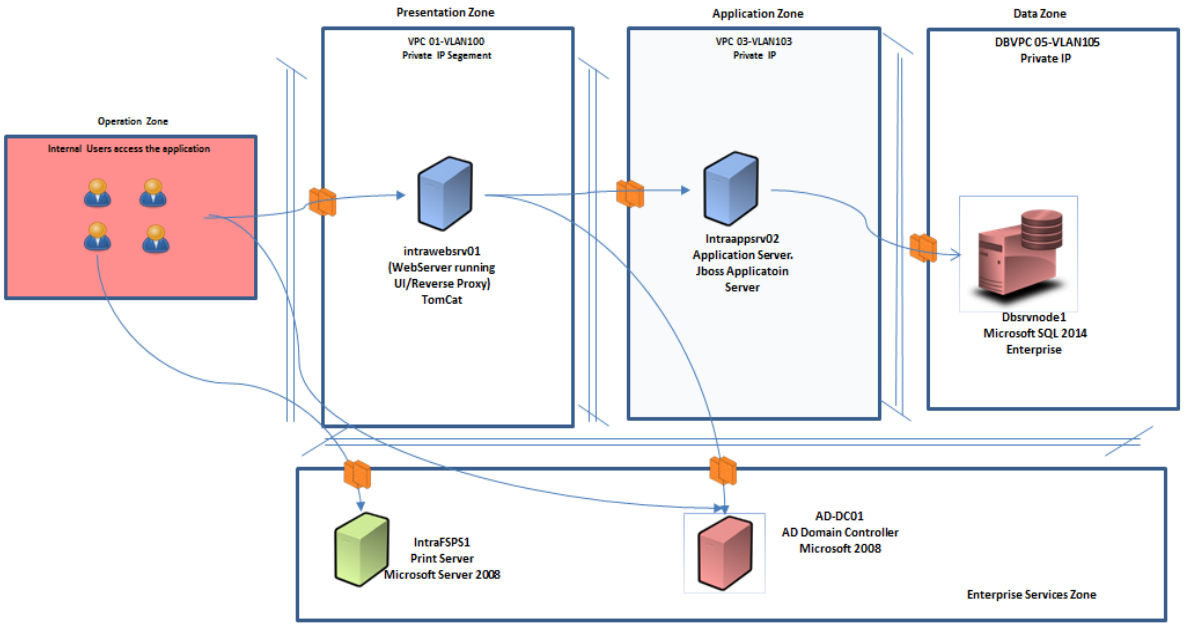


Figure 5-9 Medium Size Enterprise Architecture

The above deployment shows an internal-facing employee portal being used by a mid-size company with 35–40 employees (including seasonal workers and interns). The application uses an integrated Active Directory authentication for authentication and authorisation. The employees use the portal for information about what is happening in the company as well as to collaborate on various projects. The application also functions as the front end for marketing data analysts to perform analysis work on client sales data that are used to provide sales/promotion-related reports to clients. The portal is integrated with a file server where project documents are stored. This Java application uses a front-end Tomcat reverse proxy with user interface integrated to Active Directory authentication. The web server communicates with the application server where the WAR files are deployed which contain the code for the Java application and supported by back-end Microsoft SQL, where metadata and data produced by some of the internal transactions are stored. Table 5-4 presents the technical specification for the firm's server. In order to monitor disk usage for all servers, the monitoring tool

SolarWinds was utilised to record daily data changes of the following servers for the month of August 2016:

Server Name	OS	Processor	RAM/GB	Disk/GB
Web Server	Red Hat Linux (ESX VM)	16vCPU	32 GB	Disk 1: 80 GB- OS partition Disk 2: 80GB - Application Partition.
Jboss Application Server:(ESX VM)	Microsoft Server 2008	16vCPU	64 GB	Disk 1 : 80 GB - OS partition Disk 2: 300 GB - Application Partition
Microsoft SQL Server (Baremetal Server)	Microsoft Server 2008	32vCPU	64 GB	Disk 1: 80 GB- OS Partition Disk 2: 3000 GB - Data Partition
Microsoft Active Directory Server (Bare metal Server)	Microsoft Server 2008	16vCPU	32 GB	Disk 1: 120 GB- OS partition Disk 2: 120 GB Application Partition
Microsoft File and Print Server (ESX VM)	Microsoft Server 2008	16vCPU	32 GB	Disk 1 : 80 GB - OS partition Disk 2: 500 GB - User file repository Disk 3: 300 GB - Application File repository Disk 4: 300 GB - Application File repository Disk 5: 300 GB - Application File repository.

Table 5-4 Servers Technical Specification

Analysis of Experiment 3

From Table 5-5, it is clear that the storage requirements for data changes are less than 1% of the total hard disk for the web server, app server, and SQL database server. It is also only 1.4% for the active directory server and 1.07% for the file server.

#	Server Role	HD Size GB	Data Changes (GB)/Month	Data changes as % of HDD
1	Web Server	80	0.65	0.81
2	App Server	300	2.7	0.86
3	SQL DB	3000	7.8	0.26
4	AD	120	1.8	1.4
5	File Server	1400	15	1.07

Table 5-5 Data changes as % of HDD

VM Role		HD Size	Data Changes (GB)			Reconstructed Forensic Image after One Month (Cloud FAAS)	Daily Multiple Complete Snapshots For one month	Cloud FAAS vs Traditional as %
			Daily	Weekly	Monthly			
Server	Web Server	80	0.02	0.14	0.6	80.60	2,400	3.3
	App Server	300	0.09	0.63	2.7	302.67	900	33.63
	SQL DB	3000	0.26	1.82	7.8	3,007.8	90,000	3.34
	Active Directory Server	120	0.06	0.40	1.8	121.8	3,600	3.38
	Microsoft File and Print Server	1400	0.50	3.5	15	1,415	42,450	3.3
Total		4,900	0.93	6.49	27.9	4,928	139,350	3.5

Table 5-6 storage required in (GB) for FAAS data acquisition vs traditional

Table 5-6 shows the size of the disk changes in GB over the 30 days of this experiment. It indicates that the total amount of storage required to store all of these changes is 27.9 GB over the 30 days. Contrasting this acquiring and storing of multiple complete snapshots to obtain forensic images, this approach significantly reduces the volume of data to be stored. In the event where daily forensic data monitoring is mission critical and uses a traditional forensic approach, sustaining such storage requirements will become cumbersome. For example, having a disk forensically imaged daily for all servers would result in 139,350 GB of uncompressed space across 30 days. However, utilising Cloud FAAS requires a total of 4,928 GB (3.5% of the traditional forensic storage requirements).

5.4 Cloud FAAS Cost Benefit Analyses

The operation cost of Cloud FAAS might be a concern. Thus, it is vital to clearly outweigh the benefits of Cloud FAAS to the associated cost and this is where cost-benefit analysis is useful. To obtain a rough quotation, this section estimates a monthly bill if the organisation hosts

Cloud FAAS with a cloud provider. The Amazon Web Service Simple Monthly Calculator was used, as it provides resizable compute capacity in the cloud. This example utilises Amazon EC2 instances and Amazon EBS volumes for storage (Amazon Web Services, 2016).

In order to run the Cloud FAAS, four systems have to be configured and implemented. These are a database management system, a Cloud FAAS core system, a reconstruction system and a forensic analysis workstation.

Utilising scenario 3, presented in Experiment 3 for a medium-sized enterprise, the monthly forensic storage will be less than 5 Terabytes (TB) in order to reconstruct the forensic hard disk image for any VM at any given time during 1 month. Taking into account further storage for extra data captured from different resources, including network traffic, log systems and volatile data, it can be expected that extra data will require no more than 2 TB per month. Thus, the total storage required to host Cloud FAAS per month is up to 7 TB. Table 5-7 shows the specification suggested for the database management system and Cloud FAAS core system. Using the Amazon Web Services Simple Monthly Calculator for two high-specification instances and storage will cost approximately $1876 per month.

Fixed Cost		
System Name	Monthly Cost	Specification
Database Management System	$ 1,535.60	DB Instance 1 DB Engine is MySQL Storage 7 T
Cloud FAAS Core System	$340.38	16 GB RAM 4 vCPUs 64-bit High Network Performance- 2 GB Bandwidth
Total cost for 1 month	$1876.00	

Table 5-7 Fixed Cost for 1 month

A reconstruction system and forensic analysis workstation will run on a case-by-case basis. Utilising reconstruct engine developed for Cloud FAAS, it can be stated that reconstruction image occurs at a rate of approximately 1 GB per minute. This means that 1 TB takes around 16 hours to be reconstructed. If the analysis stage consumes the same period of reconstructing, then the complete process, including reconstruction and analysis, will cost approximately $ 144 as shown in Table 5-8.

Variable Cost			
System Name	Hourly Cost	Case Cost	Specification
Image Reconstruction System	$0.976	16 (Hour) * 5 Tera (5 servers) *$0.976 = $72.0	32 GB Memory 4 vCPUs
Forensic Analysis Workstation	$0.976	16 (Hour) * 5 Tera (5 servers) *$0.976 = $72.0	32 GB Memory 4 vCPUs
Average total cost/ per case	$144.00		

Table 5-8 Variable Cost

It can be said that the total cost of conducting a forensic image utilising Cloud FAAS for a medium-sized enterprise will cost $ 2020 per month.

Contrasting the cost of storage with the cost of performing a forensic and e-discovery obligation, the results are startling. Several studies show that 1 GB of storage costs about 20 cents to buy, but the same gigabyte will cost from $3,500 to $30,000 to conduct a review and a forensic investigation (Gonsowski, 2012; Degnan, 2011). Consistent, relentless and sophisticated cyber-attacks are causing real damage to organisations around the world, with the number of affected enterprises on the rise (Purba, 2016). For instance, Code space's Amazon AWS account was compromised and most of the company's data, backups, machine configurations and offsite backups were completely deleted. Successively, this cloud-based attack leaves Code Spaces unable to operate and went out of the business (CSA, 2016). This proves that the response to security incidents and ever more complex challenge. Therefore,

being well prepared, before an incident occurs, is most likely to lead to a successful and more cost-effective digital forensic investigation.

It is evident that conducting a digital investigation without forensics by design is an expensive process. For example, Fannie Mae, The Federal National Mortgage Association, spent approximately 9% of its total annual budget of $6 million on the discovery of electronically stored information for litigation purposes and still failed to meet its discovery deadlines. It is therefore imperative that a cost–benefit analysis is performed to ensure the proposed framework has a solid economic basis.

It is envisaged that two scenarios can be considered to implement the Cloud FAAS system. The first scenario focuses on the performance. In such a scenario, the image and backup of the VMs are taken and initially stored on the same cloud platform, and these are then transferred over to the other storage facility on another cloud, hence the data transfer rate is expected to be much faster than the data transfer rate from the cloud to the client site (Pichan, Lazarescu & Soh, 2015). The drawback in this scenario is that it is not completely free from cloud provider dependency. Knowing that up to 87% of analysis time can be saved when utilising the cloud-based forensic workflow management and processing in comparison with traditional methods (Wen, Man, Le, & Shi, 2013)

In the second scenario, the focus is on the running costs rather than the performance. The costs relating to VMs and broadband provision are contributing factors to the total monthly cost. At this point, it is worth pointing out that money spent on a high-performance server and a storage server as a one-time investment which is owned by the organisation. In this scenario, the image and backup of the VMs are captured by the agent and transferred to the storage facility on the client's site. It would take longer to download these images and carry out backups to the client's site storage facility, but it ensures that the system is independent of the CSP and, in the long

run, also offers a money-saving option as there is no monthly fee. Ultimately, it is up to individual companies to consider when evaluating whether to host Cloud FAAS on-premises or move it to trusted cloud.

5.5 Discussion

In normal forensic investigations, it is only possible to acquire an image of the system post incident. However, experiments results shows that investigators would have the ability to obtain a forensic image of the system using Cloud FAAS after, just prior to as well as hours before the incident. Therefore, this approach can not only create images that are forensically sound, but also provide access to deleted and, more importantly, overwritten files – which current computer forensic practices are unable to achieve. This results in an increased level of visibility for the forensic investigator and removes any limitations that data carving and fragmentation may introduce.

Experiment 1 shows that Cloud FAAS is able to successfully reconstruct the forensic image of the disk for any specific system at any specified time, ensuring the same data integrity as digital investigators already experience in computer-based forensics.

Experiments 1 and 2 agree that overheads related to memory are negligible, as they are less than 10 MB over a typical hour, where large volumes of disk activity/change are occurring.

However, the experiments agree that the CPU utilisation across the same period shows the agent has a larger impact. In order to reduce the potential performance impact brought on by Cloud FAAS, a number of options have been outlined, as follows:

- The non-volatile agent could be optimised to take advantage of idle system time to minimise the performance impact. However, by aligning the monitoring schedule to fit

the system utilisation monitors, triggers can be set to run when a configurable length of idle time is detected. Furthermore, the agent can be modified to consume a determinate amount of resources depending on the priority of an organisation's data acquisition.

- Increasing the time windows between clusters monitoring will reduce the potential performance impact of the agent on system resources. The agent can be designed to check for data changes at relatively wide intervals (e.g. every 30 to 60 minutes); in this scenario, it was at every 2 minutes for experimental purposes.
- The agent may also be configured to listen for 'triggers', which can be tied to specific files that are regarded as most sensitive to the organisation, such as the registry, log files and vital documents.
- The non-volatile agent could be modified to compress data before sending it to the Cloud FAAS storage. Furthermore, the agent could be enhanced to split data, and fragments could be sent depending on detected network activity.
- The agent could also be configured to have local storage on each machine for these images to be transferred to the forensic server during off-peak times. Thus, the server can distribute the work throughout the day and avoid consuming the VM's resources during busy periods.

All these solutions can also be combined to achieve an overall smaller impact of the developed agents while maintaining forensics requirements. As the size of forensic image storage might be a concern, solely recording file system changes reduces the volume of data that needs to be communicated and stored, as shown in all of the experiments. In contrast, the Cloud FAAS acquisition and storage of multiple complete snapshots or forensic images significantly reduces the volume of data to be stored. In the event where daily forensic data monitoring is mission critical, sustaining such storage requirements will become cumbersome and challenging. It is also envisaged that the Cloud FAAS need not store this data indefinitely – as the need to

investigate an incident is likely to occur within hours or days of it taking place, rather than months. As such, periodic reacquisition of the complete hard drive can take place, thereby permitting the removal of prior records from the database. This would assist in ensuring the Cloud FAAS database remains manageable. Further investigation of the cluster changes that occur on disk highlights that a large overhead exists within the file system metadata – so while a small file change might only impact a single cluster in terms of its actual content, a number of metadata files also have to be updated, resulting in, on average, eight clusters that are required to be imaged and stored. In some situations, there is an argument to suggest that monitoring only the files themselves, rather than both the files and the file system, is necessary. This would result in a significant reduction in the overall volume of data changes, with a subsequent knock-on effect to the bandwidth, processing and storage overheads. While the resulting image would not be as forensically sound, it could be demonstrated that at a cluster and file level it is.

Overall, this chapter has validated a number of the core underlying research questions that contribute to the viability of the approach.

6 Cloud FAAS Architecture and Prototype

This chapter discusses the architectural specifications of the proposed forensic acquisition and analysis mechanism in the cloud environment (Cloud FAAS) and proceeds to present a functional prototype of the system. The proposed system seeks to provide the customer as well as the investigator a fundamentally different approach to forensics acquisition and analysis within an IaaS service model which can meet business, legal and technical requirements.

6.1 Introduction

In today's world of "always-on" technology, cyber-attacks are no longer a matter of "if" but "when and how" (Kessel, 2014). Thus, a prevention strategy alone is not an option. Due to the dynamic nature of the cloud infrastructure, organisations may not be able to control forensic examination when incidents occur. It can be said that the issue of the dependence on the CSP and its associated problems are not resolved yet based upon the analysis presented in Chapter 4. The need for a trustworthy cloud forensics solution that is designed to be completely independent of the cloud service provider – requiring no intervention is becoming more apparent. Implementing such system reduces the complexity of the acquisition process, the requirement for CSPs active involvement and any modification to the CSP underlying architecture.

The sections that follow introduce an architecture for this novel approach and discuss the design and development of a functional prototype.

6.2 The Cloud FAAS Requirements

The traditional forensic process is essentially a post-event recovery of digital evidence. However, conducting this kind of procedure faces serious issues when evidence resides in the

cloud environment as the literature pinpointed. Providing digital evidence once required becomes a business requirement even before an incident occurs. Furthermore, it is imperative organisations remain in control of their data and have the ability to undertake incident analysis/forensics examination of their systems when deemed necessary and in a timely fashion. Failing to do so, may result in direct impacts on the business including damage to a brands' reputation and resulting loss of business, which are much harder to calculate.

Beside this, there are also legal and technical requirements that must be met when evidence required for incident investigation otherwise, customers may incur large fines and be subject to civil lawsuits including criminal charges (CSA, 2016). Cloud FAAS architecture is designed in a way that ensures legal and technical requirements are satisfied as follows:

- Legal Requirements

In order to succeed in a legal process, the proposed architecture should provide and protect all evidence in a manner that ensures the appropriate chain of custody and the integrity of data. Evidence acquired by the system must be logged and labeled with the identity of everyone who handles it at any stages of the investigation beginning with data acquisition and ending with the reporting. Every single action made by any user upon this evidence needs to be time stamped automatically by the system avoiding any human errors that occur during the chain of custody.

Data integrity is to ensure that evidence has not been altered in unauthorised manner. Thus, Cloud FAAS must include the forensics hashing of all image data (at all levels of data object – complete images to clusters) since the time it was acquired throughout the acquisition, storage, reconstruction, analysis and reporting with the aim of ensuring chain of custody and data integrity is maintained.

Data retention is a vital part for meeting legal and business requirements. While it is common for an organization to establish its own data retention requirements, there are certain data retention laws that must be adhered to (Posey & Burton, 2015). Different types of systems require different length of retention. In order to determine the retention time, the policy defines by the Cloud FAAS should thoroughly address legal and privacy concerns against the expense of storage. This should balance investigation and the volume of storage required in the Cloud FAAS. As the system obliges organisations to investigate an incident in a time manner, thus the retention time frame will in the region of months rather than years.

- Technical Requirements

Cloud FAAS needs to be designed in a manner that provides a richer and more complete set of admissible evidence than what current CSPs provide. Such a system should remove the need for involving CSPs when desired evidence is needed and the problems that may result from this involvement/dependence. The system should be architected to conduct a robust capturing mechanism and acquisition approach that is able to acquire the data which exists in different levels from low-level block data through the log data and to network data. The system also needs to have an ability to store, record and reconstruct an image of any specific system at any given time to the same gold standard hashing capability as experienced in traditional computer forensics. In order to ensure the confidentiality and integrity of the data in transit, all communication needs to be undertaken in a cryptographically secure manner.

Furthermore, a robust investigative access and authentication control are vital for the architecture to incorporate in order to prevent unauthorised people from having unauthorised access to the system. The system needs to define the roles and the responsibilities to be a singed to the forensic team.

Along with monitoring the data, the system does need to monitor the user interaction upon that data. A stringent audit and control system should take place from acquisition, reconstruction and analysis to the reporting stage enabling that every user act including user administrator, forensic lead and the forensic analyst are monitored. At any point in time, it should be simple to investigate who access specific data at given time then constructs the chronological line of events when needed. This leads to the ability to successfully manage any conflict of interest potentially between the members of the forensic team and manage the chain of custody.

All investigative processes need to be integrated and managed within one system. Thus, the lead investigator, for instance, can manage the complete case tracking from the point of image reconstruction through analysis and to the reporting process. All ought to be performed within a single system.

The acquisition policy is an essential component as it will directly impact its efficiency and financial cost. Thus, it is imperative to predefine a coherent policy to be incorporated in the Cloud FAAS model. Such a policy needs to be easy and adjustable enough in order directly reflect the organisational risk assessment and achieve the best trade-off between cost and granularity of data.

It is evident from the literature review that a solution to permit forensics analysis of multiple systems within the cloud has become essential. To date, most studies have worked at the techniques and tools involving one source of digital evidence at a time for investigation. However, this is not sufficient in the proposed system as there are multi-sources of evidence including unlimited VMs, network traffic, and volatile data. This highlights a need to develop new techniques and approaches which can support the interpretation of a variety of digital evidence sources. Cloud FAAS architecture needs to provide a correlation engine and visualisation component so that investigators can understand the relationship and data flow

between systems – enabling a higher level of abstraction than individual system analysis would provide.

Finally, Cloud FAAS is required to be a web-based solution as such more useful for users. This would lead to a greater level of interoperability between systems, usability for users and efficiency for management, installation and maintenance.

6.3 Design and Development of IaaS Forensics Solution

An overview of the proposed architecture is presented in Figure 6-1, with an accompanying description of each of the key components. The proposed approach in this research is to enable the cloud customer to have complete control over the forensics acquisition process, ignoring the data held by the CSP. This is made possible through the implementation of an *Agent-based* approach that sits on each of the customers VMs and communicates the necessary information to a central *Cloud Forensics Acquisition and Analysis System* (Cloud FAAS).

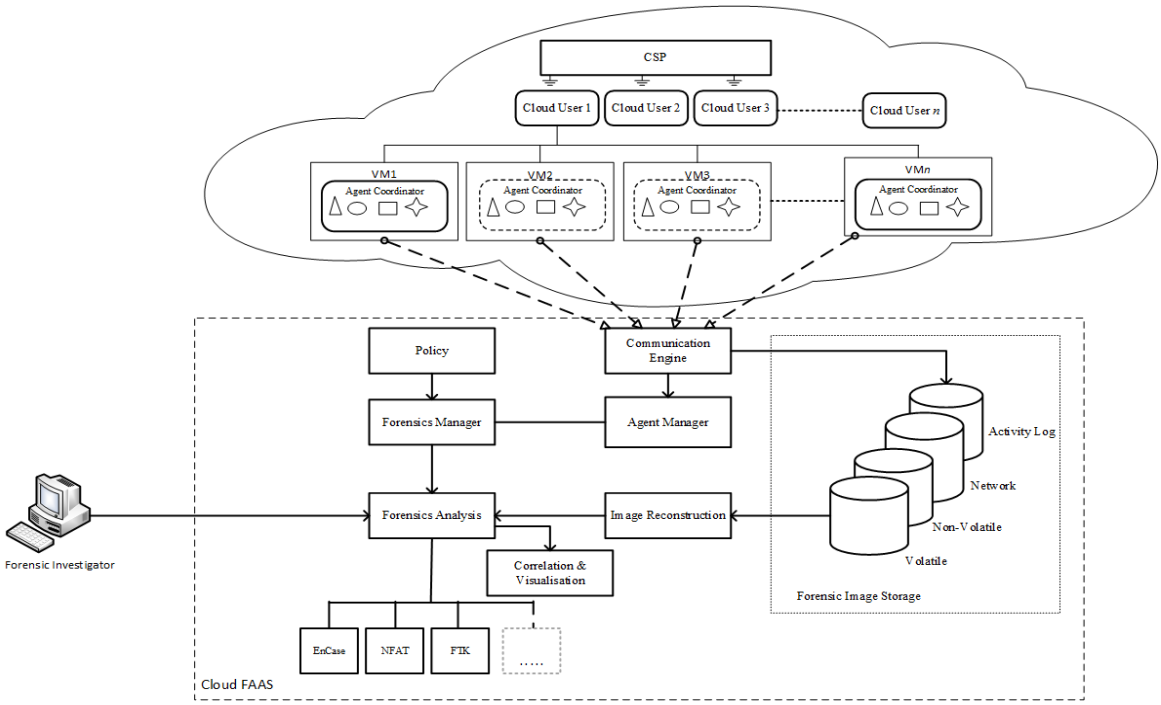


Figure 6-1 A Novel Model to Data Acquisition and Analysis within IaaS

This use of agents ensures all necessary cloud management data (e.g. VM start time/stop time) are logged. Lower level data, such as physical storage locations for the VM data, which is only accessible via a CSP, is not required due to the agent-based acquisition approach. Indeed, through the agents, it is possible to recreate an image of the VM hard drive storage at any point in time and provide access to every file in its entirety – going beyond what computer forensics is capable of achieving with partially overwritten files. This results in an increased level of visibility for the forensics investigator and removes any limitations that data carving and fragmentation may introduce.

As illustrated in Figure 6-1, the model is composed of two major components: the Agent Coordinator, and Cloud FAAS. The Agent Coordinator is responsible for the management of the agents that are installed on the individual VM system. Different agents are responsible for various aspects of the acquisition and each can be enabled or disabled depending upon *Acquisition Policy* that is defined by the cloud customer (from here on in referred to as the organisation), which is housed within the Cloud FAAS. The following agents will be available:

- Non-Volatile Memory Agent – responsible for logically imaging the virtual hard drive associated with the VM
- Volatile Memory Agent – responsible for logically imaging the live memory of the VM
- Network Traffic Agent – responsible for logging and storing network traffic (both egress and ingress)
- Activity Log Agent – acquiring system and application logs.

Which agents are necessary will depend upon organisational requirements and the nature/responsibility of the VM. For example, in a 3-tier web application, it would only be necessary to operate the Network Log Agent on the web front end system, as the back-end

server will be configured to only communicate with the web server, thus negating the need to replicate the network data store. Furthermore, the Activity Log Agent will only be required in situations where the other three agents are not in use; as such information can be derived from them. The Activity Log Agent will provide high-level log information to an investigator in environments where the overhead and cost of operating the other agents is not deemed necessary.

The Cloud FAAS is the central processing point for forensics data. It provides access to the management information, which defines the forensics acquisition policy for the VMs. The Agent Coordinator communicates with the Cloud FAAS via a Communication Engine. All necessary information including cloud management data is subsequently stored in the Forensics Image Storage.

The Forensics Manager is responsible for managing the overall system and provides the interface to the forensics investigator. It enables an investigator to select the systems to be analysed and uniquely the timeframe required of interest. The Image Reconstruction module will then take the necessary information from the image repository and reconstruct the image(s). What is reconstructed and to what data granularity will depend upon what had been defined in its policy. Having reconstructed the image, the forensics data will be sent to the forensics analysis component in order to be analysed, correlated and visualised to the investigators.

6.4 Acquisition and Data Handling

The ability to image VMs and transfer the data to a Cloud FAAS will have a huge implication for the underlying network capacity, processing overhead for the agents on the server VMs and on the Cloud FAAS infrastructure. With VM non-volatile storage in the range of 100GBs and

GBs of network activity, it could become too costly to store such data. Indeed, the introduction to a policy-based approach where different acquisition requirements are placed on different servers is an attempt to mitigate this information overload and reduce it to a manageable yet acceptable standard (forensically) for organisations. As illustrated in Figure 6-1, the data handling approach devised for this model comprises of two main steps. Initially, a forensics image of the non-volatile memory (i.e. the hard drive as seen by the VM) is taken. Operating at this logical layer on a VM means it is not possible to map this data to a physical drive – indeed; the nature of the cloud infrastructure would not permit this. As such, step 2 seeks to operate similar to an incremental backup, recording all file system changes to the drive. To ensure the forensics value of such data, the data clusters of those file changes are also stored. This allows the Image Reconstruction engine to reproduce a forensics image of the drive at any point required – including showing how files have been deleted and overwritten. Importantly, however, as the system stores these files, an investigator would also be able to obtain full access to these deleted files (something that is not possible with normal computer forensics procedures). Rather than requiring a periodic image of the complete drive, the recording of file system changes reduces the volume of data that needs to be communicated and stored.

As shown in Figure 6-2, at some point in the future it will become necessary to re-image the drive, as the volume and complexity of the file system changes that have taken place since initial imaging is such that a re-image is essential – both for computation and storage requirements. Data retention is defined by the policy but it is envisaged that this will in the region of weeks or months rather than years due to the volume of storage required in the Cloud FAAS. This approach is not devised to replace organisational backup strategies but as a means of investigating incidents in a timely fashion. It is therefore not anticipated such investigations will occur beyond a relatively short (6 month) timeframe.

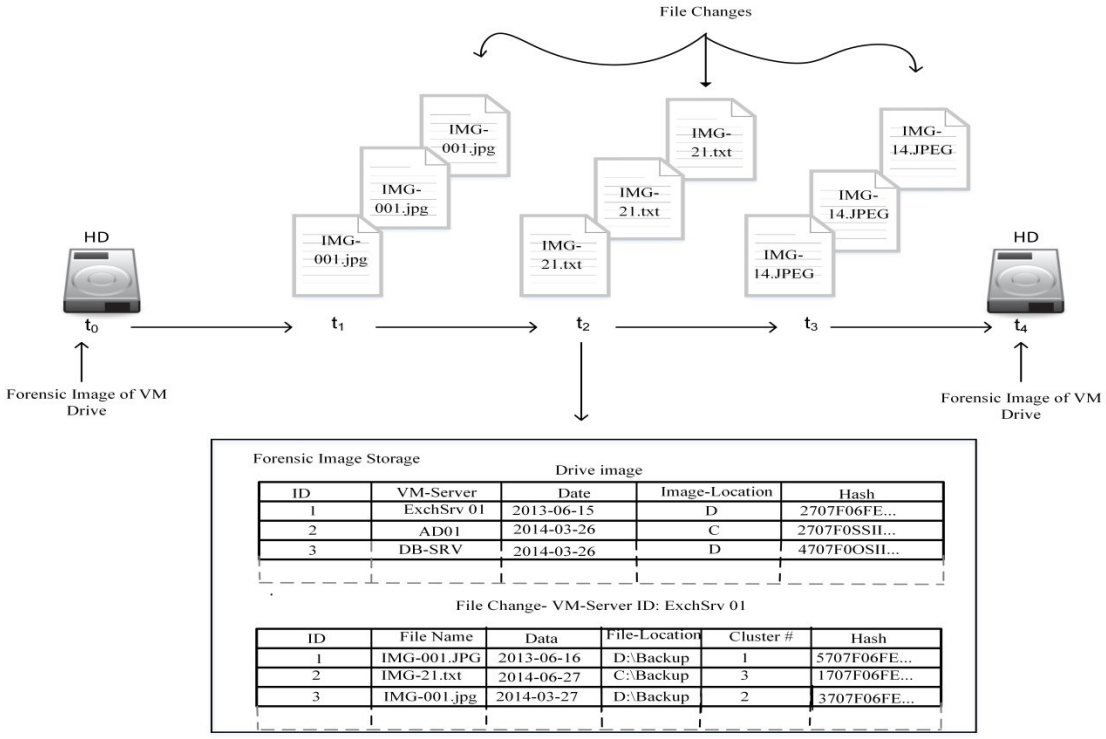


Figure 6-2 File Changes

The frequency of reimaging, the granularity of file system changes, the frequency of volatile memory captures and the resolution of the network traffic captures are all defined by the policy. Higher levels of resolution and frequency and lower granularity of data capture will all increase the demands placed upon the Agents and the Cloud FAAS – in particular, the Forensics Image Storage. Obviously, anything but the most rigorous policy will have an impact on the forensic value of the resulting data. However, this is no worse than current computer forensics – where the time acquisition takes place can have a direct impact on the quality of the resulting forensics evidence. The transmission of data from the Agent Coordinator to the Cloud FAAS can also be optimised to take advantage of low network usage to minimise any adverse effects upon the Cloud core operation.

Where a Cloud FAAS operates is largely dependent upon the organisation. They may choose to host it locally to the organisation so that ownership and access to data are as strong and reliable as possible. It could also be hosted within the same CSP as its operational servers – running as a cloud service in its own right. Whilst this is advantageous from a network bandwidth perspective – taking advantage of high bandwidth local area connections, it suffers from a single point of failure should a catastrophic incident occur to the CSP. A Cloud-based deployment more generally would certainly be advantageous from a data processing and forensics analysis perspective. Both of these aspects are computationally very intensive, yet unpredictable as to when they will be required. An elastic and flexible computing environment would allow for this – whether that is a public or private cloud.

6.5 Acquisition Policy

In order to be considered digital forensics capable, it is vital for an organisation to establish a set of policies in place that manage people, infrastructure and investigations (Almarzooqi & Jones, 2016). This section establishes an acquisition policy with the aim of maximising Cloud FAAS's ability to collect admissible digital evidence while minimising the operating costs. Each organisation will perform a risk assessment for potential loss and threats to assets and identifies associated data acquisition requirements.

The acquisition policy incorporated in the Cloud FAAS model is an essential component of the approach that will directly impact its efficiency and financial cost. To achieve the best trade-off between cost and granularity of data, the policy is based upon a set of standard templates that are derived from server roles – with critical systems having an acquisition policy that monitors all changes across all agents. Less critical systems will incorporate a less granular acquisition approach. Should an organisation desire to, it is possible to modify the template to individualise the policy to directly reflect the organisational risk assessment.

6.5.1 Policy Definition

It is vital to understand how Cloud FAAS acquisition policy works in general and identify the main factors need to be considered with the aim of deciding the level of monitoring required to meet specific forensic requirements at cost effective manner. To aid this process, organisations may wish to consider the value of assets and the consequences/impact of loss or breaches. Organisations are responsible for classifying and assessing their data and clearly state which assets are forensically important to them. Data classification helps organisations to determine and assign relative values to the data they possess; as such achieving best trade-off between cost and granularity of data. The worth of an asset can be estimated as the impact in terms of availability, integrity, and confidentiality (CIA).

Local regulations need to be thoroughly followed by organisation in order to meet specific requirements and take on different definitions based on the industry in which they are applying the policies. When it comes to information technology and security, the cost of not complying with local regulations can be significantly high in terms of fines and time invested following up on a security breach. Companies that fulfil the regulations set forth by a regulators are more likely to survive a potential investigation than companies that neglect regulatory compliance (thrivenetworks, 2016).

In cloud FAAS, the high risk level requires a greater level of resolution and frequency and lower granularity whereas low level requires lower level of resolution and frequencies and higher granularity. Regardless of what the level of system, Cloud FAAS should store it in a forensically acceptable manner where required. However, the policy is adjustable in order to directly reflect the organisational risk assessment.

Predefined options need to be identified and considered in order to keep the storage and network transmission costs to a minimum. Such elements include what agents should be activated, when they are activated and the frequency in which the images are captured and transmitted. A variety of other factors will be taken into account including data retention timeframe, metadata inclusion/exclusion and the location where Cloud FAAS resides. However, organisations may choose to make modifications to the recommended templates as per their needs. The following section discusses different factors need to be considered when creating data acquisition policies for Cloud FAAS.

6.5.2 Cloud FAAS Agents

All type of potential digital evidence resides in and can be acquired by one of the following agents:

- Non-Volatile Memory Agent – responsible for logically imaging the virtual hard drive associated to the VM and keep monitoring data changes to the disk. A large volume of evidence resides on the disk, which makes the disk an important piece to acquire; as such the agent can be configured to monitor the data changes at specific intervals. Furthermore, depending on the forensic requirements, the agent can be modified to monitor only sensitive files such as documents, registry, log files and monitor them at appropriate time periods.
- Volatile Memory Agent – responsible for logically imaging the live memory of the VM. This is concerned with monitoring volatile media which stores working files, and loses information when the file is closed before saved on the virtual hard disk – for instance in the case of a disruption in electrical power. Cloud FAAS can be configured to image the RAM at intervals, and as RAM storage is typically less than that of a storage drive, imaging RAM will be faster.

- Network Traffic Agent – responsible for logging and storing network traffic (both egress and ingress).
- Activity Log Agent – acquiring system and application logs.

6.5.3 Collection frequency

The frequency of image collection depends on policy declarations which are subject to organizational, and other regulatory requirements. For the purposes of this section, collection types are split in two:

- Continuous: The agent will monitor all data continuously.
- Scheduled: The agent can be configured to make monitor data changes of relevant targets at intervals such as every hour or once a day – depending on requirement specifications.

6.5.4 Time of monitoring (Peak vs off peak)

It is imperative that efforts are made to minimise the impact of Cloud FAAS as much as possible. Two different times are available to monitor activity:

- Peak periods: when resource utilisation is at their highest such as a disk, RAM, and processor. Depending on forensic requirements, the agent may be configured to run when there is a high activity level to enable effective change monitoring. However, some organisations might focus on performance at the expense of forensics. Thus, they utilise agents when there is a low activity.
- Off peak periods: when resource utilisation is at its lowest in order to reduce operational impact, Cloud FAAS may be configured to run when low activity levels are detected, as such reduce network activity for image transfers.

6.5.5 Data Retention Time Frame

Certain climes require data to be stored for a minimum required period of time. For example some countries require financial institutions to keep account information for a minimum of 7 years after the transaction for anti-money laundering and 30 years for anti-terrorism investigations (Metropolitan Police, 2014). A loss of such data could result in heavy monetary fines. Due to the volume of storage required in the Cloud FAAS, data retention defined by the policy will in the region of weeks or months rather than years as follows:

- 1 month
- 3 months
- 6 months

While it is assumed that these time windows will satisfy most requirements, the policies can be modified to suit individual needs if the need arises for shorter or longer time frames.

6.5.6 Cloud FAAS Storage

Cloud FAAS storage is largely dependent upon the organisation. It could be on-site so that ownership and access to data is as strong and reliable as possible. It could also be hosted remotely within the trusted CSP different from CSP hosts its operational servers – running as a cloud service in its own right. A cloud-based deployment more generally would certainly be advantageous from a data processing and forensics analysis perspective. However, the risk of suffering from a single point of failure should a catastrophic incident occur to the CSP still matters and should carefully reflect on the organisational risk assessment, especially with data that has been deemed critical to business operations (Donnell, 2016). The policy is flexible to provide both options to the organisation.

6.5.7 Metadata

Metadata is often described as “data about data” and is used to provide information about a specific file or document. The Cloud FAAS approach is monitoring the data changes at cluster level. However, further investigation of the cluster changes that occur on disk highlights that a large overhead exists within the file system metadata – so whilst a small file change might only impact a single cluster in terms of its actual content, a number of meta-data files also have to be updated, resulting in on average 8 clusters that require to be imaged and stored as shown and discussed in Chapter 5. Monitoring only the files themselves and ignoring the metadata file system would reduce the overall overhead including storage, processing and the bandwidth. The drawback of this filtering process is that the resulting image would not be a forensically sound but it is at cluster level.

6.5.8 Scenario-Based Acquisition Policy

A given organisation would like to implement Cloud FAAS in order to monitor their cloud infrastructure (IaaS) including its DB. The DB server in this scenario stores highly important information where sensitive customers, employees and financial data are stored. Thus, it is imperative that the data on it is very near to or within the time window being investigated.

A compromising of that data puts the organisation in very vulnerable position. In the event of a security incident, identifying who accessed what resource, or when a specific query was executed could prove instrumental to obtaining conclusive evidence from an investigation point of view. The data stored in the DB is considered mission critical as it could provide clues to uncovering the roots of an incident. It is important that events are treated with the highest level of sensitivity.

Furthermore, it is assumed that there are local laws pertaining to the storage and retrieval of the data. Depending on the local law, some organisations are required to provide any data requested on demand, and a failure to do so could result in unpleasant outcomes for a company, ranging from minor embarrassment to a total loss of goodwill, and multiple legal issues.

Table 6-1 is a sample selection of criteria for policy making, and will influence the policy decided on for this scenario. The volatile and non-volatile agents are continuously monitored at on and off peak periods as requests could occur at any time. Both devices – non-volatile and volatile memory- are storage media where most data resides; as such it is important that the agents monitor them at all times.

Logging of traffic anomalies provides key markers in the event of a security incident. The NIC (Network Interface Card) is monitored for network activity to enable monitoring of traffic spikes, and also requests only from specific IPs to reduce impact on operational processes which could trigger other actions based on policy specifications.

The data on the server could make a major difference in the course of an investigation. Therefore, the maximum possible period has been selected. In order to provide evidence as a means of investigating incidents in a timely fashion, data retention is set to 6 months due to the volume of storage required in the Cloud FAAS.

In order to enable high availability levels so that ownership and access to data is as strong and reliable as possible, the policy recommends locating Cloud FAAS storage locally (on site). In addition, it is assumed that the highest levels of security are needed for the data and it is imperative that the highest level of physical access controls is put in place necessitating the need to keep the data on the site.

All clusters are included to avoid any fall in the forensic standards. Despite the fact excluding the $MFT will reduce the overall volume of data changes (with a subsequent knock on to the bandwidth, processing and storage overheads), the resulting image would not be a forensically sound.

Utilising Cloud FAAS for sensitive systems will result in a much lower storage overhead while maintaining forensics requirements as evidenced in the experiment three in the previous chapter.

<table>
<tr><th colspan="5">Cloud FAAS Acquisition Policy</th></tr>
<tr><td rowspan="6">Agent Type & Frequency of Collection</td><td rowspan="2">Agent type</td><td colspan="2">Performance Opt.</td><td rowspan="2">Collection frequency</td></tr>
<tr><td>Peak Hr.</td><td>Off Peak Hr.</td></tr>
<tr><td>Non - Volatile Agent</td><td>Enabled</td><td>Enabled</td><td>• Continuous
o Scheduled</td></tr>
<tr><td>Volatile Agent</td><td>Enabled</td><td>Enabled</td><td>• Continuous
o Scheduled</td></tr>
<tr><td>Activity Log Agent</td><td>Enabled</td><td>Enabled</td><td>• Continuous
o scheduled</td></tr>
<tr><td>Network Agent</td><td>Disable</td><td>Enabled</td><td>• Continuous:
• Protocol/Port
o Network card Interface
o Scheduled</td></tr>
<tr><td>Data Retention:
o 1 Month
o 3 Months
• 6 Months</td><td>Meta Data:
• Included
o Excluded</td><td colspan="2">Storage:
• On-premises
o Cloud</td></tr>
</table>

Table 6-1 Cloud FAAS Acquisition Policy

6.6 Analysis, Visualisation and Correlation

One of the ways to manage the volume challenge would be the correlation of different sources of forensic evidence. This could provide an investigator with an holistic view of all the events

across the digital evidence sources which can be very valuable during an investigation (Casey & Schatz, 2011).

Correlation is the process to identify the links between events collected from different sources aiming at affording more timely and useful analysis. Despite the fact that there is a lack of published research in the combination of visualisation and data analysis techniques, utilising a visualisation layer above the data analysis processing can enhance an investigators understanding of the dataset and highlight how the links between data nodes have been established (Osborne, Turnbull, & Slay, 2010).

Data mining and data fusion would be employed at the analysis stage in order to perform classification, correlations and link analysis. This helps Cloud FAAS to provide digital investigation with tremendous benefits including improving the quality of decisions, reducing human processing time and reducing monetary costs. Furthermore, data mining functionality such as classification and clustering would be advised in order to extract useful patterns amongst data where the dimensionality, complexity or volume of data is extremely large for manual analysis.

After creating reconstructed images, the forensics data will be sent to the analysis component to undergo forensics examination and analysis. As indicated in Figure 6-1, after the correlation, this analysis will utilise industry de facto tools such as EnCase or FTK where appropriate. Furthermore, it provides a correlation engine and visualisation component so that investigators can understand the relationship and data flows between systems – enabling a higher level of abstraction than individual system analysis would provide.

6.7 Cloud FAAS System Security

It is vitally important to run the Cloud FAAS infrastructure in a secure fashion. The communication between the agents and Cloud FAAS must be secured. To do so, the authentication, authorisation and accountability (AAA) aspects of Cloud FAAS are considered with the aim of maintaining the chain of custody and meeting privacy and security requirements.

A stringent controlling of access to Cloud FAAS resources, enforcing data acquisition policies, auditing usages and providing the information necessary are presented as follows:

Managing the user identity is a compulsory requirement for Cloud FAAS. Before getting access to any component of Cloud FAAS, users must confirm their identity. Cloud FAAS devices a robust authentication mechanism that is not weaker than 2-factor authentication for instance, a strong password and hardware token.

Furthermore, it is highly recommended that all data passed between the cloud FAAS and the Agent coordinator remain private and integral. Thus, SSL (Secure Sockets layer) is devised for establishing an encrypted channel.

Following authentication, a user must gain authorisation to access the data portions and to perform certain tasks. Cloud FAAS employs a robust role based access control in order to ensure that access to the system is performed by authorised users. In order to maintain the chain of custody and meet security and privacy requirements, clearly separated roles are defined as per the users' job requirements, user assigns to specific roles and grants permissions as follows:

1- AAA functionality

This task manages authentication functionalities including identity management usernames and passwords, enforcing stronger passwords policy, establishing different roles and different type of access. This responsibility is assigned to and managed by the Cloud FAAS Administrator. Once Cloud FAAS administrator logged in, they can create appropriate roles and permission authorisations. Cloud FAAS accounts/roles are including security analyst, lead investigators, examiner, case reviewer and auditors. Every role is clearly defined and not overlapped with others as illustrated in Table 6-3.

2- Cloud FAAS operations

This task is for setting up the Cloud FAAS infrastructure including the database system, a Cloud FAAS core system, a reconstruction system and forensic analysis workstations. It is also to maintain a high level of performance for the whole system, for instance, performing needed tasks to ensure DB is running healthily thereby regularly checking metadata concerning the DB including the size of DB, specifying how often the data is changing, what the data retention policy is and whether to archive information. The aforementioned tasks are to be performed by Cloud FAAS Admin. However, from privacy and security perspective, Cloud FAAS Admin has no ability to read or write on the DB.

Security Analyst would be responsible for agent's configuration (install, modify, and remove), security assessment and resources utilisation. Security analysts also enforce the data acquisition policy across all VMs exist in the system.

3- Investigative tasks: The purpose of this task is to ensure that a digital investigation proceeds smoothly within Cloud FAAS and all steps of investigation – beginning with image reconstruction through analysis to reporting are examined, analysed, documented reviewed and woven together to establish a clear picture of events relating to incidents. This task is given to three different roles namely: lead investigator, forensic

examiner and forensic reviewer. Lead investigator is responsible for all levels of digital investigation – leading a digital forensic team including forensic examiner and forensic reviewer, monitoring all cases to ensuring they are assessed and prioritised in line with the delivery plan. His duties to create new case, allocate a case to specific examiner for examination, allocate examined the case to reviewer, report and attend to court when required. Also, dropping cases, archiving images, deleting images are duties of lead investigator. Forensic examiner duties are specifically to extract evidence from reconstructed images and inform lead investigator, utilising proprietary and bespoke software with responsibility for ensuring the continuity of all exhibits and maintaining integrity throughout. Following evidence retrieval, reviewing the information gathered by examiner would take place by reviewer. Reviewer assesses the completeness and accuracy of extracted evidence and verifies its integrity in order to avoid confusion, criticisms or missed evidence. Ultimately, reviewer writes up technical reports in an understandable format for individuals with limited computer knowledge, whilst ensuring that full disclosure has been provided in accordance with national guidelines ensuring compliance with best practice and legislation. Forensic reviewer and lead investigators are both have ability to attend court when required to explain the evidential investigation processes in order to confirm the validity of the evidence found.

4- Auditing task. Reviewing accountability logs in order to ensure chain of custody have been maintained and all users who access system do their tasks in a proper way. Every action performed by any user including Cloud FAAS Administrator, Lead investigator and Examiner is going to be tracked as such Cloud FAAS is capable to investigate “who did what and when”. As this helps to avoid any conflict of interest might occur between Cloud FAAS members. This task is given to Auditor role.

Task		Role					
		Cloud FAAS Admin	Security Analyst	Lead Investigator	Examiner	Case Reviewer	Auditor
AAA	Set Usernames and Passwords	√					
	Set different roles	√					
	Establish password policy	√	√				
Cloud FAAS Operations	Configure Agents install, update and replace		√				
	Maintain DB metadata	√					
	Hashing the passwords		√				
	Security Assessment		√				
	Enforce Acquisition policy		√				
Investigative Tasks	Create a case			√			
	Reconstruct Images			√			
	Examine an Image				√		
	Review a case					√	
	Report a case			√			
	Attend court			√		√	
	Drop a case			√			
	Delete an image			√			
	Archive an image			√			
Auditing	System logs – monitor all users activities						√

Table 6-2 Cloud FAAS responsibilities and Roles

6.8 Technical Evaluation

Following the development of the Forensics data Acquisition and Analysis (FAAS) prototype, a complete system will be evaluated against the following issues:

1- Performance Overhead: collecting the relevant material from different systems (VMs) across all agents and sending them to the Cloud FAAS causes the additional overhead. Adding such a new mechanism can slow down running VMs and then affect running services. Furthermore, all communication is undertaken in a cryptographically secure manner – to ensure the confidentiality and integrity of the data in transit. The *Agent Coordinator* and *Agent Manager* also include the forensics hashing of all image data (at all levels of data object – complete images to files) to ensure chain of custody and

data integrity is maintained throughout the acquisition phase. This will require some techniques such as encryption and hashing which adds more workload to the running services as well.

2- Network Overhead: it is vital to keep network overhead to a minimum. However, the volume of data being recorded – VMs data and Network activities- and sent via network activities resulted in an impact on the network traffic.

3- Granularity: it refers to the level of data which is being collected and stored. Although the finer level of integrity is important to digital investigators, it can affect the level of performance overhead.

4- The quality of Reconstructed image: the constructed image retrieved from *Forensics Image Storage* must gain acceptance from both the judicial and technical communities.

5- Performance of Forensics Analysis: some cases are high profile, such as a child abduction, which must be processed as quickly as possible in order to provide investigators with time-sensitive information that may be vital to the outcome of the situation. Unfortunately, some of these cases can take hours or even days to finish on larger evidence. The average amount of data per case, as experienced by FBI's 15 Regional Computer Forensics Laboratories, has grown 6.65 times (from 84 GB to 559 GB) in eight years (2003–2011). Thus, it is imperative to reduce the overall processing time of large quantities of data by leveraging the power of a high-performance computing platform and adapting existing tools to operate within this environment.

Ultimately, in order to mitigate the aforementioned issues and gain a better result, the best balance between the various issues has to be taken into account.

6.9 Cloud FAAS Prototype Implementation

This prototype aims to reflect how Cloud FAAS deploys a non-volatile agent, reconstruction engine and validation images. Figure 6-3 shows the main web interface for the Cloud FAAS users.

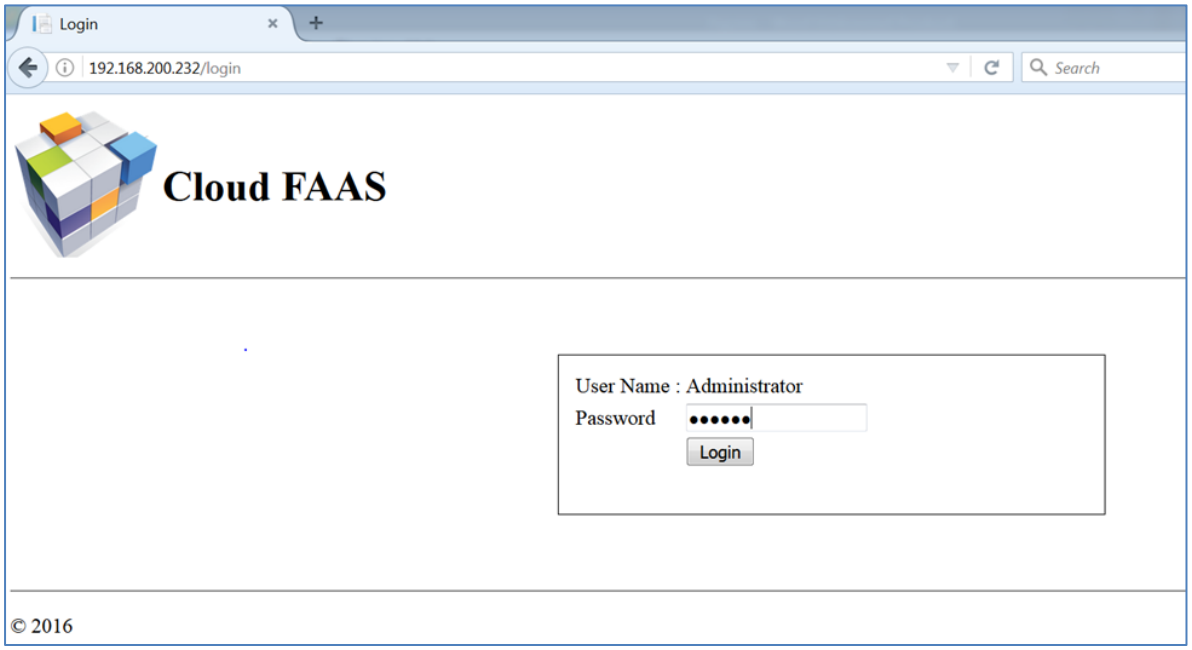


Figure 6-3 Cloud FAAS Main Interface

Every user has a specific task to conduct, for example, the System Administrator who has full system access including administrative tasks. Such tasks include agent's installation on any system, the policy definition and facilitating the granting of multiple privileges or roles to users. Once the user is authenticated, all running systems would show up. Figure 6-4 shows a range of available systems that are ready to be investigated. The system admin can also define the policy, specify any VM to be investigated, activate or deactivate any agents.

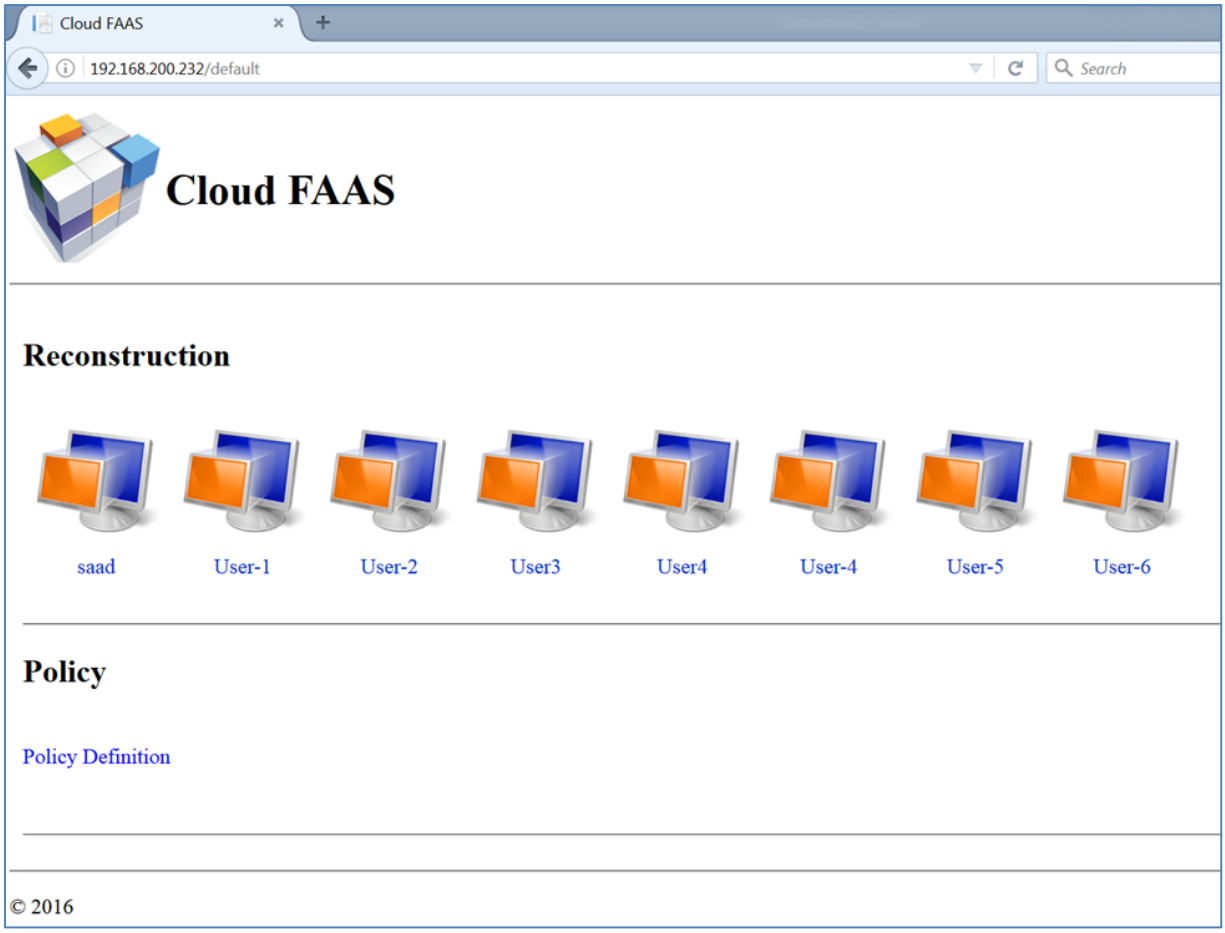


Figure 6-4 Running Systems

Once the desired system to be investigated is specified and image reconstruction is required, an investigator from the organisation will select the date/time stamp, as shown in Figure 6-5.

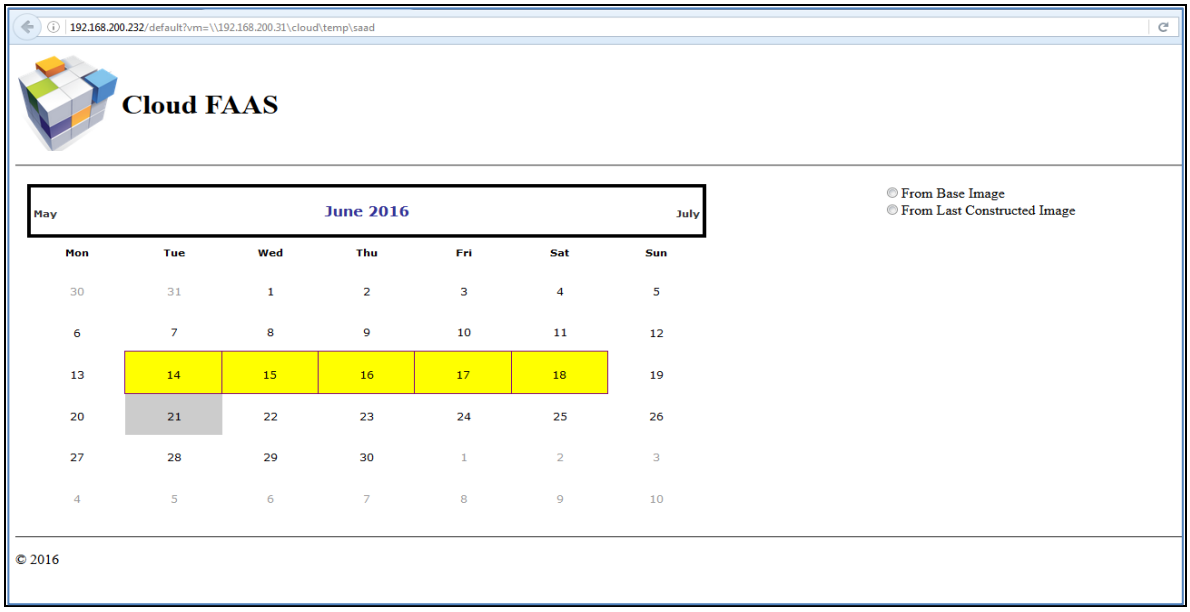


Figure 6-5 Required Date to Reconstruct Specific System

There are ranges of possible restore points labeled with their time stamps as illustrated in Figure 6-6. The frequency time set for the non-volatile agent to monitor data changes is defined by the policy acquisition as stated in section 6.5 of this chapter. In this scenario, it was at every 2 minutes for prototypal purposes.

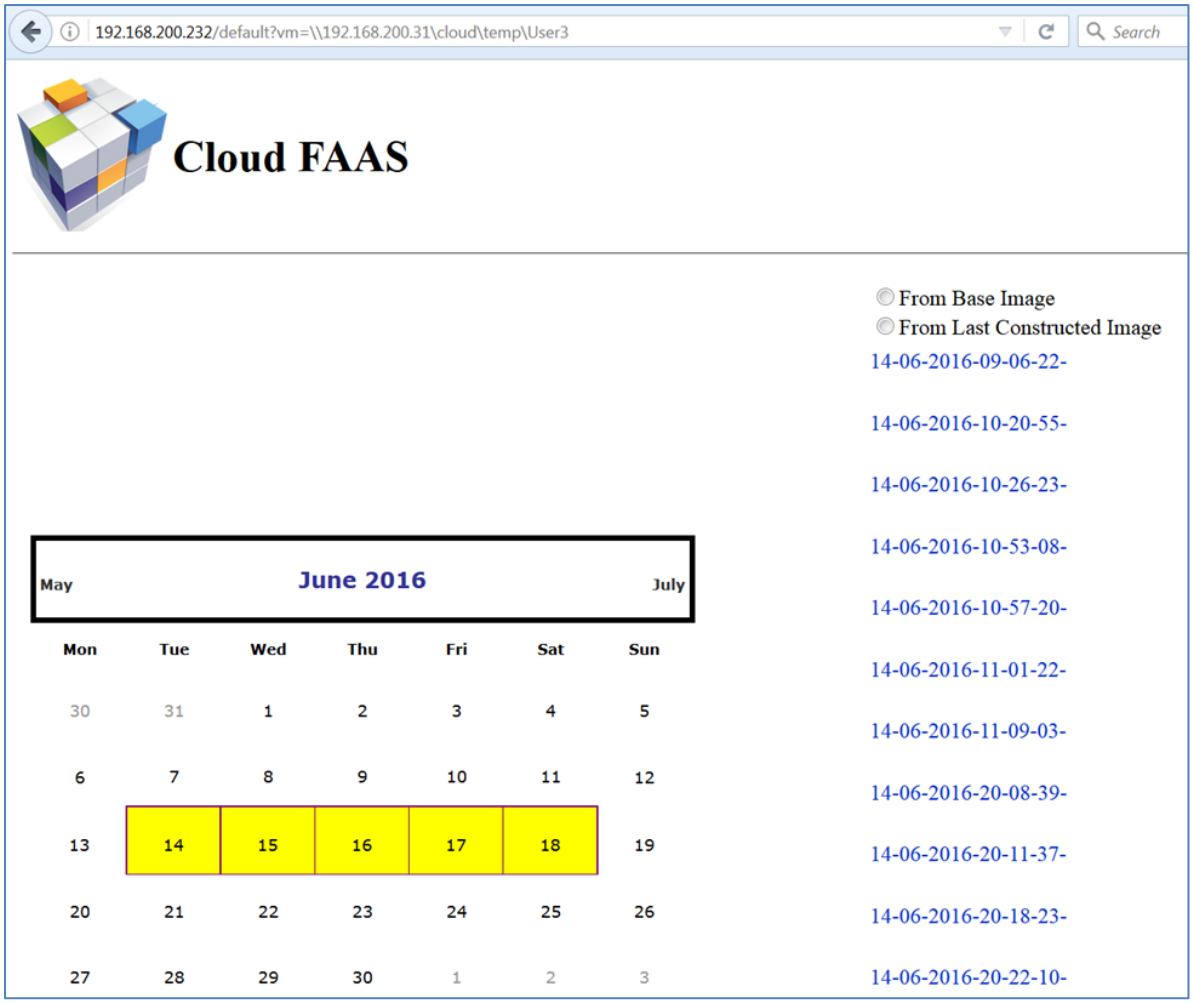


Figure 6-6 Data Changes Time Stamps

When image reconstruction is required, an investigator from the organisation will select the date/time stamp and the reconstruction engine will create the forensic image, as if it were forensically acquired as per normal (i.e. a complete bit-for-bit copy). While the desired image is reconstructing, a progress bar is a pop up in order to visualise the progression of reconstructed image as illustrated in Figure 6-7.

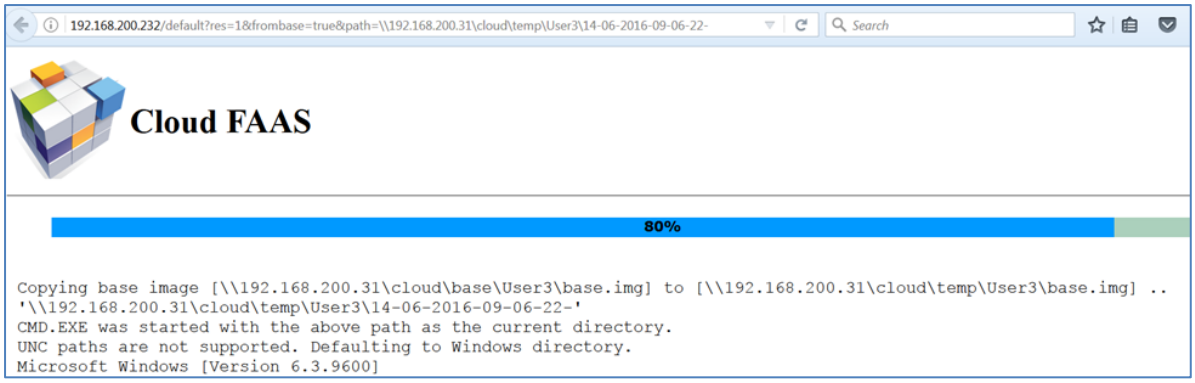


Figure 6-7 Reconstruction Progress Bar

Once the required image is constructed, the system validates the image integrity by comparing the generated hash values from the hard disk immediately after the data changes (which are already stored with data changes) against the generated hash value of the reconstructed images, ensuring the same data integrity as digital investigators already experience in computer-based forensics as shown in Figure 6-8.

Figure 6-8 Hash Validation

When the forensic examiner access the system, h/she can only access the reconstructed images specified by the forensic manager as shown in Figure 6-9.

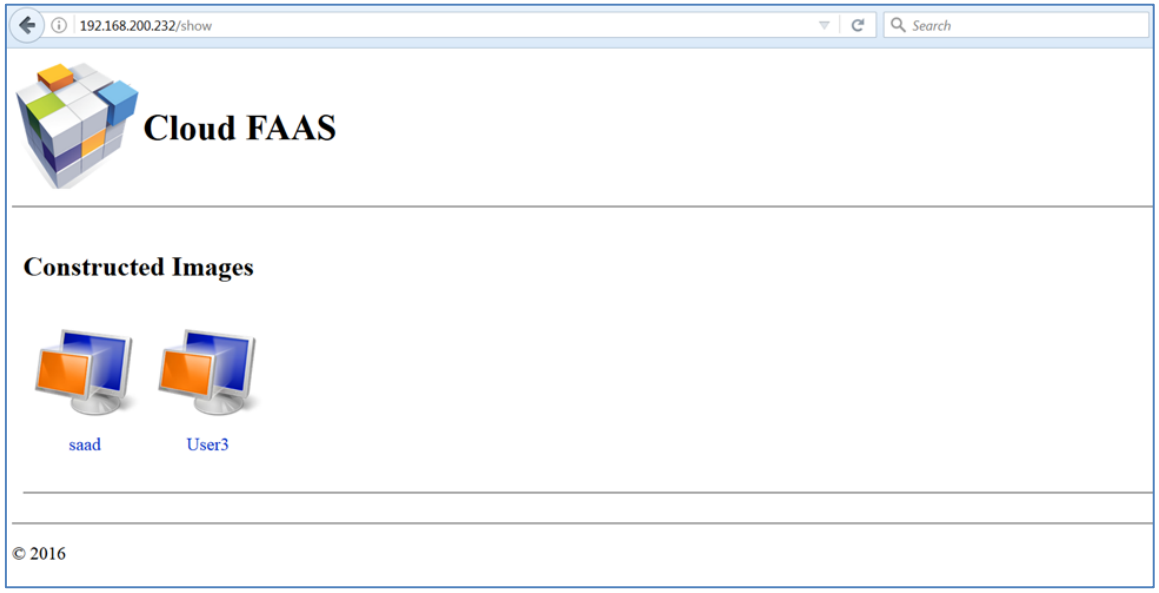


Figure 6-9 Reconstructed Systems

The forensic examiner has the ability to select any reconstructed system at a given time to be analysed as illustrated in Figure 6-10.

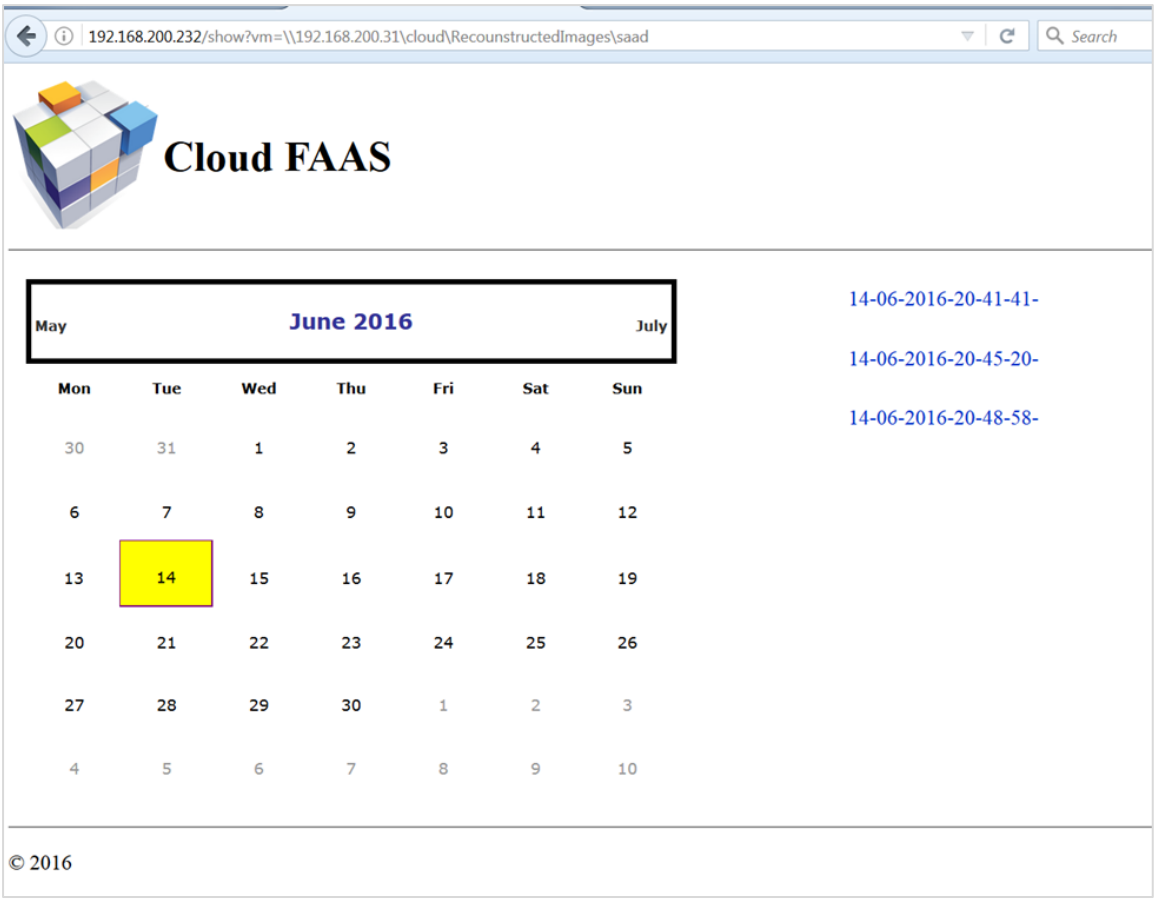


Figure 6-10 Reconstructed System Time Stamp

The forensic examiner can download the image in his forensic workstation for further investigation and analysis as shown Figure 6-11.

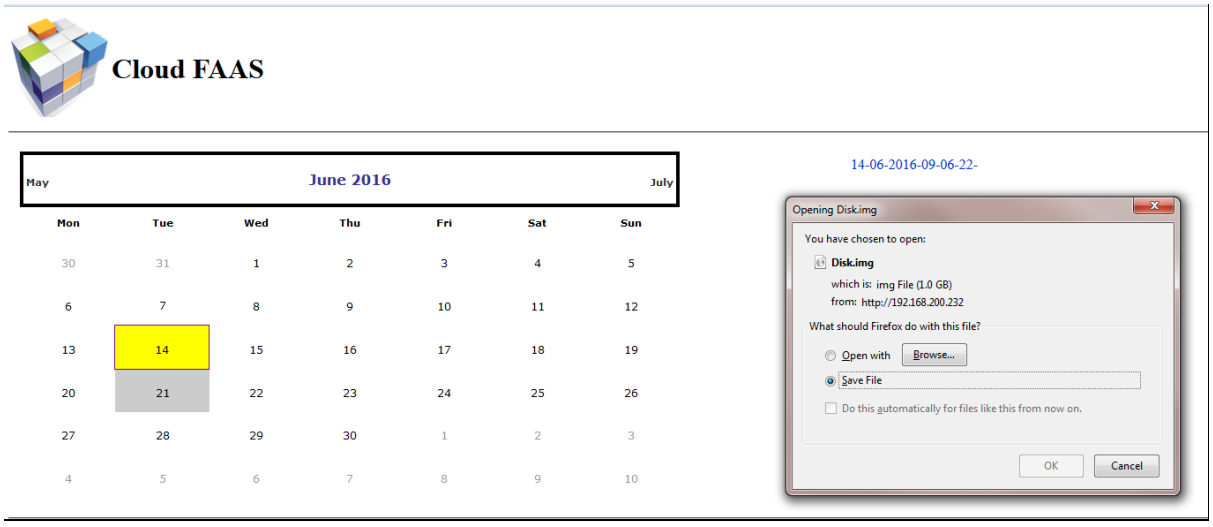


Figure 6-11 Downloading of Reconstructed Image

The reconstructed images are uploaded to and examined by forensic tools including free forensic tool kit such as Autopsy 4.0.0 as illustrated in Figure 6-12.

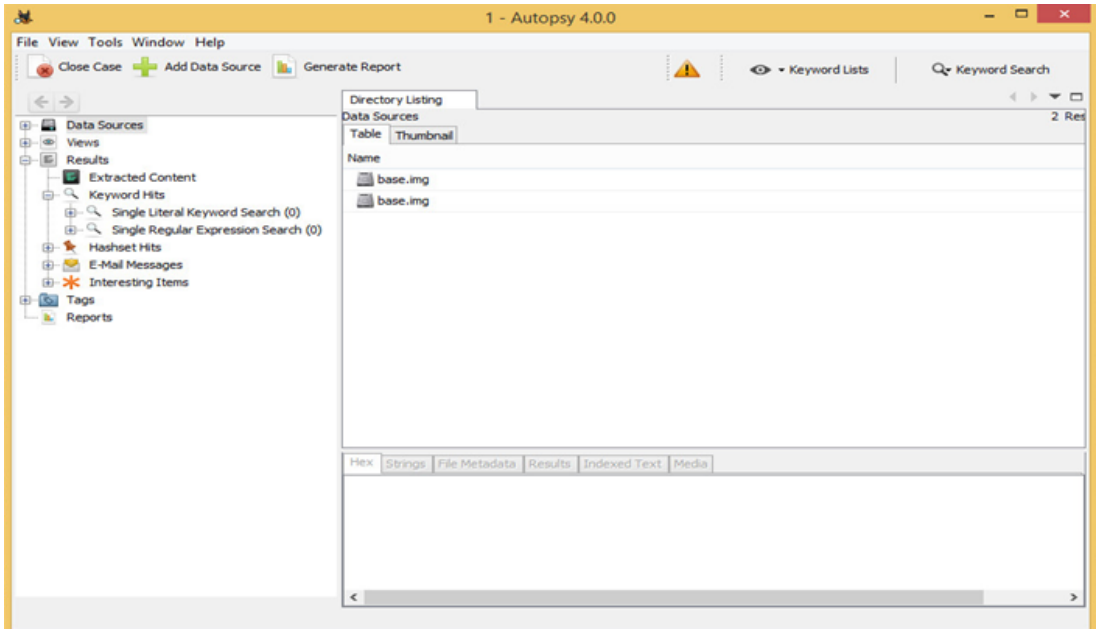


Figure 6-12 Uploading Reconstructed Images to Forensic Toolkit (Autopsy) for Analysis

Once the case is thoroughly analysed, the forensic examiner will send the report back to the forensic manager as shown in Figure 6-13.

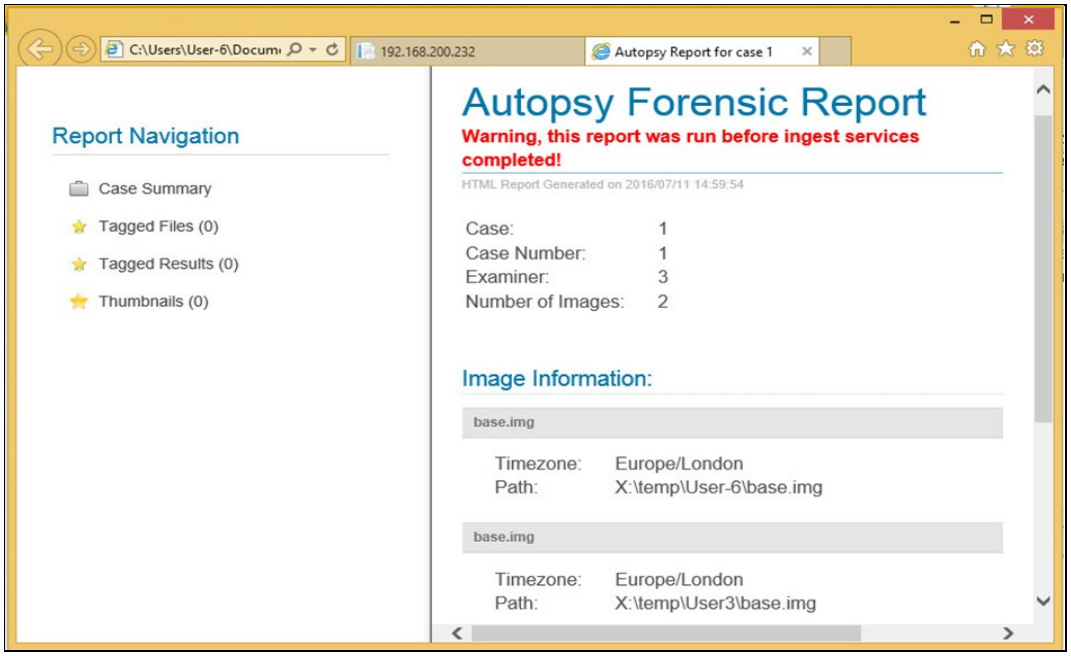


Figure 6-13 Reporting

6.10 Conclusion

The architectural specifications of Cloud FAAS have been designed in a modular and robust manner to enable the customer as well as the investigator a fundamentally different approach to forensics acquisition and analysis within an IaaS service model. The mechanism has been designed on the principle that reduces the complexity of the acquisition process, the requirement for the CSP's active involvement and any modification to the CSP's underlying architecture. Cloud FAAS is designed to store, record and reconstruct an image of any specific cloud based system at any given time to the same gold standard hashing capability as experienced in traditional computer forensics. Thus, organisations remain in control of their data and have the ability to undertake incident analysis/forensics examination of their systems when deemed necessary and in a timely fashion.

The proposed architecture meets both legal and technical forensic requirements with the aim of maintaining the chain of custody, securing the system, monitoring the data as well as people who access this data.

The acquisition policy has been established in a fashion to enable organisations the flexibility and convenience of predefined composite factors in order to meet their specific requirements, reflect the organisational risk assessment. The architectural model enabled all investigative processes to be integrated and managed within one system. Thus, a complete case can be tracked from the point of image reconstruction through analysis and to the reporting process. All performed with in a single system. Based upon this architecture, a functional prototype has been presented.

7 Evaluation of Cloud FAAS

In addition to the promising validation results obtained from experimental packages, there is a need for additional qualitative evaluation. The objective of this evaluation is to provide quantitative feedback received from professionals in different areas including academics and industry practitioners in order to support the results performed in experimental work. For the qualitative evaluation, the research literature states that the ranges typically include up to ten persons: 6 persons (Creswell, 2007), 6-8 persons (Kuzel, 1999) and 6-10 persons (Morse, 2000). In order to cover academics and practitioners point of views, 12 experts have been contributed to this evaluation including academics and forensics practitioners. The participants, who were all experts on the subject matter, were carefully selected and a set of questions was carefully designed and presented to the experts. This was followed by a detailed interview and discussion with the experts on the different aspects of approach, which was covered in the open-ended questions. Finally, the chapter discusses the findings of the evaluation along with the results towards the end.

7.1 Introduction

An expert-based evaluation has taken place with the aim of appraising the performance and identifying the limitations cloud FAAS. People with an academic background are important in this evaluation due to the educational context of the research and the opportunity to receive feedback that is adequate for Ph.D. research. However, the viewpoint of forensics practitioners is also crucial due to the context of the research and its practical orientation.

Different points of view from experts with different backgrounds and experiences help to cover all dimensions of the offered transdisciplinary research. The experts who evaluate cloud FAAS

are from different countries. They have different experience and knowledge background in different areas.

A number of the interviews were conducted remotely, and interviewees were located in divergent countries/jurisdictions, as such presenting different legal and technical point of views. Professionals from different areas of academia, and industry including cloud forensics experts, cyber security experts, police officers, and forensics investigators comprise the candidates. This diversity will prove instrumental to a further room of development in the area and explore the strengths along with the weaknesses from different perspectives. The responses have been grouped and analysed by subject areas.

Prior to the interview, a brief video was sent to the experts in order to familiarise them with the research problem, the model architecture and the experimental results. The video is about 18 minutes long and available at https://www.youtube.com/watch?v=7h2gbxmdvFI

The questions were designed to obtain perspectives from different areas to create a high level picture of the research area, and its industry applications. The questions were asked face-to-face, emails or Skype interviews.

7.2 Interviewees

Experts represent the main scientific areas and practical themes that were included in the offered research including cloud forensics, cloud security and digital investigation. The search process of the experts was realised via the Internet as follows:

- Members of committees thematically related to the research area of scientific conferences.
- Authors of work thematically related to research articles in scientific journals.

- Scientists from related fields working also as lecturers and/or as administrative staff members in an educational institution.
- Forensic practitioners in the field of cyber security, digital investigation and digital forensics.

The following list of the complete experts' details provides more information about participant's experiences and their research interests. The experts can be categorised into two groups either academics (6 people) or practitioner (6 people).

- Academics

  A1- Sameera Almulla, Ph.D. Sameera is a research assistant at Khalifa University of Science, Technology and Research, United Arab Emirates. She gained a Ph.D. in digital forensics of cloud computing at Khalifa University. She also gained MSc in digital forensics with distinction at Khalifa University. Samerah had an internship in digital forensics and information security at Royal Holloway, university of London. She published several papers and articles specifically in cloud forensics.

  A2- Ameer Al-Nemrat, Ph.D. Ameer is a senior lecturer at the School of Architecture, Computing and Engineering, University of East London (UEL). He also the Director of Professional Doctorate in Information Security & the MSc Information Security and Digital Forensics programs. In addition, Ameer is the founder and director of the Electronic Evidence Laboratory, UEL. He is an active researcher in the area of cybercrime and digital forensics where he has been publishing research papers in peer-reviewed conferences and internationally reputed journals. He is a co-editor of the book "Issues in Cybercrime, Security, and Digital Forensics". He also was the guest editor of the special issue of the International Journal of Electronic Security and Digital Forensics (IJESDF). Ameer

has led a Cybercrime Programme Project with a German institution, which won the "Good Practice Award from The European Commission under the Leonardo da Vinci scheme which focuses on the teaching and training needs of those involved in vocational education and training.

A3- Manabu Hirano, PhD. Associate Professor at National Institute of Technology, Toyota College, Japan. He is a visiting senior lecturer at the university of Kent UK. Manabu worked for Toshiba Corporation in the mobile Internet applications group. He has a certification of Information system security administrator approved by Ministry of Economy, Trade, and Industry, Japan. He has published several articles and research papers in information security and digital forensics.

A4- Hussain Aljahdali, Ph.D. He has been an academic researcher at King Abdulaziz City for Science and Technology (KACST) since 2007. KACST has joined the International Council for Science (ICSU) as a National Member. He gained a PhD in cloud computing security from the University of Leeds. Hussain is a research active in the area of cloud security where he has been publishing research papers in peer-reviewed conferences.

A5- Richard Overill, Ph.D. FBCS, CITP, CEng, FIMA, CMath, CSci, FHEA. He has been a senior lecturer (Associate professor) in Computer Science King's College London, Department of Informatics since 1975. Dr Richard has published several papers in the field of digital forensics and has presented at various international conferences. He is an active researcher in the area of information assurance and information warfare; cybercrime forensics and statistics; anomaly detection and intrusion detection.

A6- Alshareef Mohsen, Ph.D. He gained a Ph.D. in Computer Technology from the Florida Institute of Technology, USA. He worked for Ministry of Interior- Public

Security at Kingdom of Saudi Arabia. He held the position of IT Department Director, Data Centre Director, and Computer Crime Unit Director. He has also been involved in imparting training related to digital forensics to various personnel ranging from police officers to law enforcement. Mohsen has more than 20 year of experience as a professional information technology manager. He joined ACITS in January 2011 with full responsibility for ACITS as a CEO for the Company. He has directed and supervised IT professionals and staff in public as well as private sector companies.

- Practitioners

P1- Sibt Ali is currently working in Metropolitan Police as a digital forensic specialist. He is responsible for the forensic acquisition, examination and analysis of evidence from a variety of electronic based systems and devices to assist in the prosecution of offenders. He worked with West Yorkshire Police as a Hi-Tech forensic officer from 2009 to 2015. Sibt gained his BSc in Computer Science from the University of Hull in 2004.

P2- Eman badri was a Senior Hi Tech Officer at West Yorkshire Police-UK for more than 10 years. Eman currently is a digital forensic consultant at UNICEF. She gained her MSc degree in computer forensics from the University of Bradford in 2005 and Bachelor's degree in computer Software Engineering from the University of Hull in 2004.

P3- Asif Iqbal is information security expert and digital forensics investigator and researcher with over 17 years of experience in diverse technical, senior management advisory and consultancy positions. His research interests are including Cybercrime investigation, digital forensic Investigation cyber warfare and strategic defence. Asif has published several papers in the field of digital forensics and has presented

in various international conferences. Asif has several industry recognised certifications including Certified Information Systems Security Professional (CISSP), Certified Information Systems Auditor (CISA), Certified Information Security Manager (CISM), Certified in the Governance of Enterprise IT (CGEIT), Certified in Risk and Information Systems Control (CRISC), Certified Fraud Examiner (CFE), Certified Ethical Hacker (CEH), Computer Hacking Forensic Investigator (CHFI), ISO/IEC 20000 Certified Consultant.

P4- Muhammad Irfan is currently a senior security specialist at Workplace Safety and Insurance Board (WSIB), Ontario, Canada. He was an Incident management Analyst at TNS Smart Networks, Ontario, Canada. Mohammad is Certified Information System Security Professional (CISSP), Certified Information Security Auditor (CISA) and Access Data's FTK Examiner (ACE) with over 10 years of experience in Information Security.

P5- Mano Panchartnam is the founder and CEO of CyberSentry which offers a range of Cyber Security services, all designed to help enterprises address Cyber Threats proactively. With over 15 years of IT Security experience in the Government Sector, Mano has found CyberSentry Inc. with a mission to help clients deal with all their Cyber Security challenges. Mano previously served at Ontario Ministry of Government and Consumer Services as Security Design Coordination from 2001 to 2006 and Security Design Manager from 2006 to 2014.

P6- Salah Altokhais, Cyber Security Specialist. Salah holds a master degree in Computer Security, graduated with distinction from RMIT University, Australia in 2009. He is Certified Forensic Analyst (GCFA), Network Forensic Analyst (GNFA) and CISSP. He has been working as a forensics analyst and information security researcher at KACST since 2007. He has conducted internal researches on File

Systems Forensics, Analysing Incidents Artefacts and Memory (RAM) Forensics. He is a member of the digital forensics and incident response team (DFIR) in ECP department at KACST.

7.3 Interviewees Response Evaluation

The questions were designed in a manner that investigates Cloud FAAS in terms of its validity, admissibility, efficiency, reliability and usability. An open question was raised at the end of the interview with the aim of appraising weaknesses as well as strengths of the proposed approach.

The following sections analyse feedback from each expert, and general conclusions will be drawn based on each expert's perspective: academia/professional. In addition, further analysis will be made where applicable based on specialisation of each interviewee.

In order to make comparisons and see the different opinions of each expert about the same question, each of the questions posed to the participants is discussed and analysed. This permits for a more comprehensive method of evaluation and gives a different degree of detail in answers. The following sections discuss, analyse, compare, and evaluate each participant's opinions, as revealed by their responses to the questions posed to them.

7.3.1 Thoughts on research problem

This section seeks to gain information about the expert's viewpoint on the identified research problem— to what extent do researchers and practitioners find the Cloud problematic when making forensic investigations? All the academics and practitioners agreed that the research problem was a valid.

- Academics

  Each interview started by ensuring that the interviewee had a general understanding of the research area. A6 saw it as a clear and well-defined research problem, and thought there was a gap between the problem statement and the currently available technology. A1 supported this perspective, saying Cloud FAAS would provide "a lot of value in terms of automating evidence acquisition". It is noteworthy that P3, an industry professional, cited A1's research in passing (this is included in the literature review). Other academics, including A5, described the research as "interesting". A2 referred to the rapid expansion and advancement of Cloud infrastructure, an expansion not matched by a similar advancement in forensic solutions, and described the problem as "valid". A4 agreed that the research problem was genuine; saying that as the Cloud is used in cyber warfare, to conduct forensics on such a battleground would be very challenging.

- Practitioners

  The professionals interviewed generally came from a legal and investigative background. P3 indicated an appreciation of the research area, and seems to have kept abreast of current developments in the field, as he mentioned some researchers who work in this area, including A1, whose response was discussed earlier. P2, who worked for West Yorkshire Police, stated that many Cloud-based cases have been investigated, but the process was really cumbersome when more than a few Cloud Service Providers (CSPs) were involved, especially when they were located in different jurisdictions. One common issue she placed emphasis on that investigators had to take on trust any information provided by CSPs.

P6 identified the problem and described it as a big challenge in today's Digital Forensics and Incident Response (DFIR) spectrum. P6 had personally worked on a few investigations that included CSPs; he said that he had encountered many legal and technical challenges.

P4 recognised the problem, saying that it was obviously challenging to conduct a forensics investigation in a Cloud environment, as the investigators were trying to uncover the underlying source and root cause of an incident while having no physical control, directive search or any ownership to deal with the case.

P5 agreed that performing IT forensics using client infrastructure hosted in the Cloud is indeed a challenging task—the problem statement accurately describes some of the challenges faced by some organisations. He thought that a Cloud FAAS service could provide a solution, if designed and used with care and security.

P1 believed that there is a niche application for such research. The same sentiment was echoed by other professionals, with terms such as "time critical" and "next big thing".

7.3.2 Efficiency of the Cloud FAAS approach towards data acquisition and image recovery

This section analyses feedback on the feasibility of industry applications. Issues such as cost, performance, storage, and acceptance are discussed.

All experts interviewed agreed on the need for such a solution, but many raised concerns about performance utilisation and a willingness to adopt the product by organisations. The majority of responses to this question were affirmative.

- Academics

A6 said Cloud FAAS generates a solid case for evidence since the MD5 function is used as a cryptographic hash function. A5 thought the approach sounded feasible, and others expressed similar sentiments. A4 raised concerns on the use of storage, as the Agent's continuous monitoring of disk activity could result in massive amounts of generated data. However, once he realised that the Agent applies an incremental approach to the images at cluster level, he saw the value of a method that helps to address the storage issue to a considerable extent while maintaining forensic requirements.

A1 also stated that any bid for automation in Cloud forensics would be greatly appreciated. She thought that a Cloud FAAS, with its capability to automate evidence acquisition, would really add value for a forensic research team.

- Practitioners

P2 said that a Cloud FAAS would help speed up investigations, as the traditional forensic ways of attending the data centre in person were not feasible when the centre or centres might be anywhere in the world. A Cloud FAAS makes perfect sense, as the forensics can be carried out in the Cloud base.

P6 thought that the approach and the concept were great, and provided a much-needed solution. However, he suggested a few modifications and improvements. For example, he advocated building extra, advanced features to the core functionality of the Cloud FAAS, with the idea of enabling threat hunting—the current hot topic in DFIR.

P1 believed that getting time-slice images of the hard drive would be useful, saving investigators' time, as it would help to trace user activities at any given time, removing the need to examine the entire hard drive.

P5 thought the approach was good in a broad sense, but he raised a security issue with the implementation and addressed some of the challenges that would have to be overcome before making the service available to clients. He pointed out the need for security safeguards for remote cluster capture and data transport. The captured disk images would need to be transferred with transport layer encryption, and once in forensic storage, would need to be further encrypted to prevent the CSP from having access. P5 also suggested having a form of Key Management Service within the Cloud FAAS solution, so that keys used to encrypt the captured image could be managed by the Management Console.

7.3.3 Evidence Admissibility

This question explores feedback from the experts on the admissibility of evidence extracted by Cloud FAAS. The interviewees generally agreed that such evidence would be admissible in court as the MD5 hash function is accepted forensically to prove data integrity. They saw the admissibility from a technical perspective. The majority of respondents believed that Cloud FAAS indicated that the acquired evidence is a complete and accurate copy of the data contained on the original device. It is pointed out that if the acquisition and verification hash values are matched then no alteration of the evidence can have taken place. Despite this, several highlighted that different country' jurisdictions had different definitions as to what constitutes admissible evidence.

- Academics

  A2 said that admissibility of evidence depended on many factors, including technical, legal, and Service Level Agreements in the Cloud (SLA). If we assume the existence

of law enforcement collaboration, the Cloud FAAS approach may maintain the admissibility of collected evidence.

A1 thought that the validation process added value to the acceptance of forensic images, as very well-known tools like Encase and FTK use hash techniques for integrity verification—it can therefore be asserted that this approach generates forensically sound evidence. Despite this, she recommended consulting a lawyer and expert witness to obtain an accurate answer to the question of evidence admissibility. A3 also recommended consulting a trusted third party as well as technical experts.

A6 believed that the evidence would be admissible in any court of law, as long as the approach maintained the hash value (digital fingerprint) to prove that the evidence had not been tampered with.

A4 shared this view from a technical perspective, but he argued that from a legal standpoint, each country's own legislators would decide what was or was not acceptable. A5 did not completely agree but did think that admissibility was totally dependent on local laws. This sentiment was echoed by most of the academics interviewed, who showed general uncertainty from a legal perspective.

- Practitioners

P2 thought that the use of hashing techniques, along with a clear explanation of the Cloud FAAS approach, would make the extracted evidence acceptable to the court. P4 emphasised the need to maintain the chain of custody in order to document the chronological history of the handling of extracted evidence. However, as Cloud FAAS records every single action and presents the images with integrated time stamps, it

provides an automatic chain of custody with accurate documentation free from introduced human errors.

P1 pointed out the difficulty in building an image at the specific point and time and then proving its identity with the information on the original drive, to which there is no physical access. A very clear explanation would be essential. P1 also believed that admissibility of the evidence would depend on each country's law, P6 echoed that sentiment.

P3 thought the approach to be technically correct but found it difficult to comment legally; he would like to see a little more detail about the underlying processing.

P5 pointed out the need for the Cloud FAAS to be certified frequently by independent auditing firms to ensure compliance with industry security standards. Audits would be needed to ensure that people, processes, and technology (including the underlying Cloud Facility) were all compliant to prevent the validity of the evidence being questioned in court.

7.3.4 Thoughts on the Cloud FAAS capabilities in terms of undertaking an investigation and its usability

It is vital to investigate the usability and capability of Cloud FAAS in order to assess whether organisations can utilise Cloud FAAS to forensically investigate incidents in a satisfactory matter. Thus, this question posed to have feedback from experts to evaluate Cloud FAAS capabilities and usability. The general perception is that the interface needs improvements but it is easy to deal with usability wise. Nevertheless, definitive feedback cannot be given as it is a new model, and hasn't attained a wide acceptance.

- Academics

From a technical perspective, A2 pinpointed that the cloud FAAS could add a value and contributes towards finding a comprehensive solution. A number of academics such as A6 suggested improvements to the interface and integrate it into a commercial product in order to make it more attractive to the public and such as good enough to sell in the market. However, he indicated its current interface is sufficient for a proof of concept research application as it does the job for Ph.D. work.

A4 suggested enhancements such as enabling automatic optimisation based client's behaviour. A number of academics such as A1 found it very easy to use and keeps all the headache of imaging behind the Cloud FAAS. However, integration with existing forensic tools such as FTK and Encase would be a plus for cloud FAAS.

- Practitioners

Industry professionals interviewed echoed similar sentiments to the academics. P3 and P4 indicated they were impressed with the capabilities, but indicated the need for an improved GUI, and clear documentation for end users, and researchers.

P2 indicated that it sounds fine to her especially that extracted evidence/images can be uploaded and examined via traditional tools such as Autopsy or FTK as shown in the video. P1 shared this view and pinpointed that as a forensic investigator they would be able to cope with. Furthermore, P6 appreciated the concept if it addressed the issue of the network speed, storage, maintenance and management requirements. P4 put emphases on the capability of acquiring evidence images at any given time as such correlates it with the incident time, that is very helpful for speeding up and solving in digital investigation.

P5 also pinpointed that the functionality of the Cloud FAAS solution is a viable solution for forensic and security team. However, there must a level of the concept of operation in terms of process, safeguarding the Cloud FAAS infrastructure including encryption.

7.3.5 Application to digital investigations

One of the main interesting features provided by Cloud FAAS is that the investigator would be able to acquire a forensic image for any specific system at any given time. This section analyses feedback from multiple experts as regards how the usefulness and performance enhancements such feature might provide to the field of digital investigations.

Experts interviewed were of the general opinion that it will speed up investigations, and critical data can be obtained on demand from any point in time. However, reservations were expressed as to the performance impact of FAAS.

- Academics

  There is a general consensus that having a forensic image of the system when required is critically important as investigations are often held back due to the hoops involved in acquiring evidence. A1 highlighted that it will speed up the process a lot because the current cloud based senarios, investigators need to communicate with service provider to gain such images. This process takes a long time and usually CSPs do not respond properly. A1 referred to the research conducted by (Dykstra & Sherman, 2012), when they tested forensic acquisition at the host operating system level by exercising Amazon's Export feature. This experiment most closely resembles the process probably used to satisfy subpoenas and search warrants. Using expedited shipping, it took five days to receive data, at a cost of $125. Thus, cloud FAAS is helpful in speeding up the investigation. A5 also mentioned that it will not only speed up the but also provides

more live information. A2 pinpointed that capturing windows within the virtual machine may be considered as one of the proactive solutions, which Cloud FAAS is offering as demonstrated. A1, A3, A4 and A6 are from the same perspective and see a valid application of the method, provided the performance impact is negligible.

- Practitioners

Industry professionals interviewed echoed similar sentiments to the academics and agree that having the image readily available to be investigated when required is a great concept. P2 saw its importance that cloud data usually are volatile and no longer available to access. For example, when investigator needs to investigate what happened a couple of months back, it is hardly likely that such data is been overwritten, Cloud FAAS system appears to get that data back and hence it made it easier for us as examiners.

P4 described this feature as the core of the centre of all investigation of the case. P6 also stated that it will help massively since in the nowadays situation this will take ages to discuss with the CSPs and get a physical copy delivered by mail. P1 believed that speed is essential to forensic investigations however; he argued that the ability to go through multiple images at different time windows is more important than the speed. His point of view supported that it provides more information in one way but in another way it potentially can be more time consuming to examine multiple time slice in case of uncertainty of exact time of the incident therefore, a lot of manual work needs to analyse what happened. Thus, some kind of automation analytic will help a lot.

7.3.6 Time sensitive image acquisition and relevance to feasibility levels for forensic investigators

The posed question aims to investigate the level of visibility for the forensic investigator when Cloud FASS obtains an image after, just prior to or hours before the incident. To what extent this results in an increased the level of visibility?

- Academics

  A4 sees the value of providing the image before and after the incident as this would help the forensic investigators of telling a clear and concise story as it arranges events chronologically. He added that timeline analysis becomes critical, as does determine when the incidents (malware for example) was created and/or executed on a system. This can easily identify the incidents behaviour.

  A3 is of the perspective that while speed is important, and which Cloud FAAS provides it may take a while for it to become an industry, and subsequently legal standard. A2 is also convinced FAAS will add value to current investigative solutions and stated that the credibility of such solution will be relied on to which extent integrity of data is maintained.

  A5 shares a similar sentiment; he indicated that the service will prove useful depending on the time granularity required for the particular system/case.

- Practitioners

  Practitioners echoed academic's sentiments relating to being able to get images before, during, and after an incident gives an investigator a lot of more capabilities to resolve the cases and greatly help in differential forensic analysis. P4 similarly thought that it

helps in understanding the system state before and after an incident and this will make life easier as images can be obtained from any time period. P1 also finds the ability to have multiple exhibits, while being able to select isolated time exhibits a boon to investigative capabilities.

P6 raised a concern with the speed of the acquisition especially in the case of the investigators do not know which server was compromised - which is the most case as he claimed. Thus, all systems need to be downloaded and investigators work on them one-by-one. However, parsing few artifacts and reviewing timelines helps a lot to determine which server needs to be analysed without actually downloading the full image. For example, parse the $MFT and USN Journal of every system and generate a timeline of activities for every day included with the network traffic and windows security logs. So an investigator can review this timeline to identify any suspicious files modified, created and deleted without actually downloading the full image.

7.3.7 Technical considerations, costs, and performance impacts of the Cloud FAAS on end user

This section analyses feedback as regards the impact of a running Cloud FAAS agent on end users. The question was that from the VM user perspective to have a better understanding of implications and overheads resulting from Cloud FAAS.

Experts interviewed were of the general opinion that the reliability and application of the method are beneficial enough as long as the resource impact is not operationally impacting in a noticeable manner.

- Academics

A5 thought it completely dependent on the type and level of policy. A6 places emphasis on that it is designed to run in the background and not interrupt regular processes. Thus, it will have no impact on the users as long as the VM performance was not degraded. A4 touted the ability to configure the cloud FAAS predefined policy to considering networking and computing optimisation will reduce the impact significantly on the end user side especially that the end users are not responsible for installing, maintaining or managing the agents as shown the demonstration. Cloud FAAS is being an investigative system, besides this sees its value in terms of auditing and troubleshooting tasks as well which are important for end users perspectives.

- Practitioners

P5 thought that as long as the agent does not put a load on the client VM this should not be an issue to the VM user. P2 is of the opinion that a wider acceptance of the method will kick start acceptance in a legal context. The end user needs to comply with local legislation to hold a certain amount of data for a specific amount of time (different countries have different legislation). Thus, the end user needs to compromise some computational resources for complying with local legislation.

P6 identified a cost constraint. Four agents running on a busy VM server will be costly. Despite the incremental nature of the backups addresses that to an extent, he is still concerned about performance impact in case all agents are running. P4 sees the reliance on native system resources as a cost saving attribute of the method, in addition, the incremental nature of the backups was seen as a mitigating factor as regards storage and bandwidth.

7.3.8 Predefined/Customisable policy options, and cost impact

Interviewees were of the general opinion that enabling organisation's determined their policy based upon a set of predefined standard templates that are derived from server roles with the aim of finding a balance between the overhead of the Cloud FAAS service against the organisation needs. This section explores expert views about how these predefined settings might help to alleviate costs concerns when applied by beneficiaries.

- Academics

  A6 sees the configurable policies as making it a more usable system as relevant values are not hard-coded. A4 is of a similar opinion, and described the option to configure time windows as having a "huge impact". A1 and A3 express similar sentiments as regards configurable time windows.

- Practitioners

  P2 sees the ability to select time frames for analysis as being able to save investigators time and money. Same can be said for end users. P1 indicated the ability to customise settings will also increase adoption rates as nothing is hard coded, and users are not forced to deal with one option. P3 also mentioned the policy factor "can significantly be important." Customisations were also described as a "really good feature" by P4.

7.3.9 Reliability/attainability/feasibility of Cloud FAAS at an operational level

This section analyses feedback from multiple experts as regards the mitigation effect on a real-world adoption and operational feasibility of Cloud FAAS. Interviewees were of the general opinion that it definitely addresses a problem, and will need widespread adoption to make a definitive assessment.

- Academics

  A4 believes it is feasible, as it is designed to run seamlessly; furthermore, the images generated are easy to retrieve. A2 sees it as a potential answer to some problems in the forensics research space. A5 sees it as attainable with the right amount of investments, and interest. A similar sentiment was echoed across the board, with academics generally perceiving it as an interesting research solution.

- Practitioners

  P6 feels it is feasible, and with sufficient testing, can become standardized operationally. P4 also considers it a straightforward solution and considers it attainable based on the demonstration of the method.

7.3.10 Cloud FAAS's Strengths and Weaknesses

This section analyses feedback from experts as regards the strengths and weakness of Cloud FAAS, a section outlines both, from an academic and practical point of view. Interviewees were of the general opinion that it definitely addresses a problem, brings advantages to forensic and security committees but the cost of storage and the performance impact are still challenging.

The majority of interviewees highlighted that constructing a forensic image at predetermined intervals, prior or after the incident, would be the main strength of the approach. Such key feature helps in revealing suspect activities and generating a timeline for analysis (A4 & P6), accessing to data that would takes month to get hold off or never to get it because it has been deleted or even worse overwritten (P2) automating data acquisition (A1), adding capability for differential forensic analysis (P3) and detecting the user behaviour based on tracking their activities (P1).

P4 believed the strength of the approach to be the manner in which it attempted to bridge the gap that exists in cloud technology including relying on CSP when an incident occurs. P5 pinpointed that providing predefined policies and easily modifying them is advantageous to Cloud FAAS.

A4 highlighted more two advantages of Cloud FAAS that there is no requirement for CSP involvement or modification to the CSP's underlying architecture and digital investigators do not have to have special skills of investigations as Cloud FAAS utilises common tools including FTK and EnCase. P6 underlined the importance of having centralised analysis for multiple systems that Cloud FAAS provides. He emphasised on its importance in solving cases as it explores how the links between different data nodes from different systems have been established.

As regards weaknesses, the concerns of cost and storage were raised by most of the interviewees. Furthermore, P3 would have liked to have seen more tests in large scale in order to understand the real implication of Cloud FAAS. P6 highlighted that there is the potential performance impact and extra overhead derived from maintaining and managing the running agents set on the customer VMs.

7.4 Conclusion

In addition to promising results conducted in the experimental study, it was imperative to evaluate Cloud FAAS by receiving unbiased and objective feedback from different experts with different backgrounds including academics and practitioners. The level of expertise of all the interviewees and their critical insight was very useful to establish the actual practicality, appraise the performance and identify the limitations Cloud FAAS. The questions were

designed in a manner that covers the main areas involved in this transdisciplinary research including technical, legal and business requirements.

The general outlook of the experts interviewed is one of interest, and positive. A definite need for Cloud FAAS model has been identified, and there is a considerable level of interest in the field, and varied industry applications. The majority of respondents thought Cloud FAAS addresses that gap, and provides a cost-effective, and resource friendly approach to cloud forensics. From a technical perspective, the majority of experts were positive about the admissibility of evidences acquired by Cloud FAAS. However, some experts found it difficult to judge the approach legally as it heavily depends on the local jurisdiction law. Furthermore, there is a concern as regards costs, resource usage, and acceptance legally, and in the industry. In addition, as suggested by some of the participants, the research concept could be further tested in a more sophisticated and larger scale environment in order have a better understanding of Cloud FAAS implications. There are areas which require further research and there was abundant room for improvement, despite the overall satisfactory outputs. This research would be taken further based on the inputs received from the participants, and the shortcomings needs to be resolved with the aim of enhancing the capabilities of Cloud FAAS.

8 Conclusion and Future work

The research aimed to define, design and develop a novel approach to forensic acquisition and analysis within an Infrastructure as a Service (IaaS) cloud model that both meets existing standards and shifts the control of the forensic analysis to the cloud user (who actually owns the data) rather than the cloud provider. This aids in reducing the complexity of undertaking the data acquisition process, the requirement for the CSP's active involvement and any modification to the underlying architecture of the CSP.

This aim was achieved by examining the current state of the art to define the gap need to be addressed and by carefully studying the possible and most suitable approaches to tackle the problem. Thus, extensive experiments were performed using different scenarios to validate the defined concept, and the result was evaluated by experts within the field.

8.1 Achievements of the Research

Overall, the research has achieved all the objectives initially set out in Chapter 1, with a series of experimental packages and studies undertaken for the development of Cloud FAAS. The main achievements of this research are:

1- Developing a novel approach to forensic acquisition and analysis within an Infrastructure as a Service (IaaS) cloud model. Cloud FAAS enables the IaaS cloud customer to have complete control over the forensics acquisition process, without the need to depend upon the CSP.
2- Design and development of a novel clusters-based acquisition algorithm that is able to recreate disk images to the same standard as existing computer acquisition processes.

3- Developing a serious of models that are experimentally proven to understand the relationship between various operational factors such as bandwidth, memory requirements and processing overheads.
4- Proposed a flexible acquisition policy to provide an appropriate trade-off between managing the security requirements of consumers and the associated costs.
5- Development of a functional prototype based upon the specification of Cloud FAAS, illustrating its admissibility and functionality.
6- Evaluation of the feasibility of the approach by seeking expert opinions and feedback from both practitioners and researchers.

8.2 Limitations of Research

Although the objectives of the research program have been met, a number of decisions had to be made which imposed limitations upon the work. These decisions were typically either practically based or due to time restrictions. The key limitations of the research are summarised below.

1- The focus in the final prototype has been on a non-volatile agent as it represents the most challenging aspect of the acquisition process. The remaining agents were not incorporated into the prototype due to insufficient time and resources. However, it is expected that standard existing forensic acquisition approaches would suffice (e.g. memory imaging to capture the volatile memory, TCP Dump to capture network traffic).
2- Cloud FAAS acquisitions are multi-sources of evidence including unlimited VMs, network traffic and volatile data providing an investigator a holistic view of all the events across all digital evidence sources which can be very valuable during an investigation. However, this highlights a need to develop new techniques and approaches which can support the interpretation of a variety of digital evidence sources.

Unfortunately, the lack of available time hindered the author in developing such techniques.

3- The test scenario utilised, whilst realistic for workstation based use in IaaS, does not reflect the complete range of uses for IaaS services – particularly servers. It does not provide a good basis for understanding the scalability issues that might exist. Running Cloud FAAS within larger scale environment to provide a better understanding of technical implications and to reflect real world scenarios was one of the main barriers due to the sensitivity and confidentiality of the data to be experimented upon. The limited number of virtual machines is a significant barrier to better understanding the effectiveness of the proposed approach.

8.3 Scope for Future Work

This research program has advanced the field of cloud forensic. However, a number of areas of scope for future work exist, specifically related to this research and more generally within the area of cloud security. These suggestions are listed below:

1- Designing the acquisition agents in a way that results in a minimal impact upon the core activities of the system. For example, designing the agents with functionality to specifically use unused processing cycles, optimising the frequency of capturing based on the workload of the target server especially highly utilised infrastructure such as DB, File Servers etc.

2- Deploying Cloud FAAS in wider and practical cloud environment that reflects the complete range of uses for IaaS services – particularly servers. This would permit a comprehensive and thorough evaluation of the technical implications and costs resulting from such a system on the day-to-day operation of a cloud system.

3- Having acquired the data, further research will also be directed towards the correlation and analysis of forensic images (volatile, non-volatile, many systems, network data, logs) that are often required in more complex systems and to which current tools lack.
4- Having correlated the data, utilising a visualisation layer above the data analysis processing would definitely enhance an investigators understanding of the dataset and highlight how the links between different data nods have been established.
5- Further research needs to be done with the aim of identifying which industry needs Cloud FAAS, how often organisations do forensic investigations and on what target hosts. Some industries are at higher risk of falling victim to cybercrime than others such as financial services, healthcare and governments. Therefore, the Cloud FAAS costs and pricing to organisation needs to take into consideration all these aspects and still would have to be affordable for the organisation to be utilised.
6- Cloud FAAS demonstrated the ability to reconstruct the forensic images of any given system at any given time as it was working before or after an incident. As such, Cloud FAAS has the potential to be improved to serve as a disaster recovery system that meets the goals of preserving and collecting data in a manner that is legally defensible and forensically sound.
7- Maintenance of the agents needs to be looked at. After the agents are installation, there is a need for a management dashboard to re-install, update and replace agents.

8.4 The Future of Cloud Forensics

As there are increasing cloud-computing uses, there is a growing need for trustworthy cloud forensics. It is clear that the cloud is here to stay and is growing with every passing minute. This leads to an increasing trend in illegal activities involving clouds, and the reliance on data stored in the clouds for legal proceedings. Several researchers have identified and explored the challenges confronting the digital investigators when they conduct forensic investigations in cloud-based cases. Therefore, some studies have proposed technical solutions to mitigate these challenges. However, there are still open issues that need to be tackled. Current cloud architectures do not support digital forensic investigators, nor comply with today's digital forensics procedures largely due to the dynamic nature of the cloud. Whilst much research has focused on identifying the problems that are introduced with a cloud-based system, to date there is a significant lack of research on adapting current digital forensic tools and techniques to a cloud environment.

References

Academy, B. (2012). PHYSICAL SERVERS VS. VIRTUAL MACHINES. Retrieved from http://www.backupacademy.com/blog/physical-servers-vs-virtual-machines.html

ACPO. (2012). ACPO Good Practice Guide for Digital Evidence, 1–41.

Ademu, I. O., Imafidon, C. O., & Preston, D. S. (2011). A New Approach of Digital Forensic Model for Digital Forensic Investigation. *(IJACSA) International Journal of Advanced Computer Science and Applications, 2*(12), 175–178.

Adolph, M. (2009). *Distributed Computing : Utilities , Grids & Clouds*.

Al Fahdi, M., Clarke, N. L., & Furnell, S. M. (2013). Challenges to digital forensics: A survey of researchers & practitioners attitudes and opinions. In *2013 Information Security for South Africa - Proceedings of the ISSA 2013 Conference* (pp. 1–8). http://doi.org/10.1109/ISSA.2013.6641058

Alecsandru, P., & Patriciu, V. (2014). Logging System for Cloud Computing Forensic Environments. *CONTROL ENGINEERING AND APPLIED INFORMATICS*, *16*(1), 80–88.

Alherbawi, N., Shukur, Z., & Sulaiman, R. (2013). Systematic Literature Review on Data Carving in Digital Forensic. In *Procedia Technology* (Vol. 11, pp. 86–92). Elsevier B.V. http://doi.org/10.1016/j.protcy.2013.12.165

Almarzooqi, A., & Jones, A. (2016). Chapter 3 A FRAMEWORK FOR ASSESSING THE CORE CAPABILITIES OF A DIGITAL FORENSIC ORGANIZATION (pp. 47–65). http://doi.org/10.1007/978-3-319-46279-0

Almulla, S., Iraqi, Y., & Jones, A. (2013a). A Distributed Snapshot Framework for Digital Forensics Evidence Extraction and Event Reconstruction from Cloud Environment. In *2013 IEEE 5th International Conference on Cloud Computing Technology and Science* (pp. 699–704). Ieee. http://doi.org/10.1109/CloudCom.2013.114

Almulla, S., Iraqi, Y., & Jones, A. (2013b). Cloud forensics : A research perspective Cloud forensics : A research perspective. http://doi.org/10.1109/Innovations.2013.6544395

Almulla, S., Iraqi, Y., & Jones, A. (2014). A STATE-OF-THE-ART REVIEW OF CLOUD. *2014 ADFSL*, *9*(4), 7–28.

Almulla, S., Iraqi, Y., & Jones, A. (2016). Digital Forensic of Cloud based Storage Snapshot. In *nternational conference on Innovative Computing Technology (INTECH 2016), Second International Workshop on Cloud Security and Forensics (IWCSF 2016)*. Dublin.

Amazon. (2014). AWS Import/Export - Cloud Data Transfer & Migration Services. Retrieved May 25, 2014, from http://aws.amazon.com/importexport/

Amazon Web Services. (2016). Amazon Web Service Simple Monthly Calculator. Retrieved November 2, 2016, from https://calculator.s3.amazonaws.com/index.html#r=IAD&s=EC2&key=calc-1A2BD7E2-A6FB-4C59-9900-75BB5B8C9713

Aminnezhad, A., & Dehghantanha, A. (2012). A Survey on Privacy Issues in Digital Forensics, *1*(4), 311–323.

ARUN, S., & GANESH, B. (2005). *Role of Open Source Software Development in Digital Forensic Tools*.

Aydin, M., & Jacob, J. (2013). A comparison of major issues for the development of forensics in cloud computing. In *8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)* (pp. 77–82). London: Ieee. http://doi.org/10.1109/ICITST.2013.6750166

Ayers, D. (2009). A second generation computer forensic analysis system. *Digital Investigation*, *6*, S34–S42. http://doi.org/10.1016/j.diin.2009.06.013

Banas, M. (2015). *Cloud Forensic Framework For IaaS With Support for Volatile Memory*. Retrieved from http://trap.ncirl.ie/2068/1/matusbanus.pdf

Barrett, D. (2013). Security Architecture and Forensic Awareness in Virtualized Environments. In K. Ruan (Ed.), *Cybercrime and cloud forensics* (Vol. 2013, pp. 129–131). USA: IGI Global. http://doi.org/10.4018/978-1-4666-2662-1.ch006

Biggs, S., & Vidalis, S. (2009). Cloud computing: The impact on digital forensic investigations. In *Internet Technology and Secured …*. London. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5402561

Birk, D. (2011). Technical Challenges of Forensic Investigations in Cloud Computing Environments, 1–6.

Birk, D., & Wegener, C. (2011). Technical Issues of Forensic Investigations in Cloud Computing Environments. In *2011 Sixth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering* (pp. 1–10). Okland,CA: Ieee. http://doi.org/10.1109/SADFE.2011.17

Birman, K. (2010). *We ' re still in the early days … Year 1 for cloud computing may be closer to 2015 "*. Athenas.

Brodkin, J. (2008). Gartner: Seven cloud-computing security risks. Retrieved January 27, 2015, from http://www.infoworld.com/article/2652198/security/gartner--seven-cloud-computing-security-risks.html

Brown, E. (2012). NIST Special Publication Helps to Demystify Cloud Computing. Retrieved January 4, 2014, from http://www.nist.gov/itl/cloud-052912.cfm

Business Insider. (2015). BI Intelligence projects 34 billion devices will be connected by 2020. Retrieved November 26, 2016, from http://www.businessinsider.com/bi-intelligence-34-billion-connected-devices-2020-2015-11?IR=T

Carrier, B. (2001). Defining Digital Forensic Examination and Analysis Tools Defining Digital Forensic Examination & Analysis Tools Brian Carrier.

Casey, E. (2001). *Handbook of computer crime investigation: forensic tools and technology*. Retrieved from http://books.google.com/books?hl=en&lr=&id=DQdWRitMcyAC&oi=fnd&pg=PP2&dq=Handbook+of+Computer+Crime+Investigation:+Forensic+Tools+and+Technology&ots=t4tFY_4UrQ&sig=y22a3Gi2xGoW2bTR9-_P5sHr33Y

Casey, E. (2011). *Digital Evidence and Computer Crime*. *igi-global.com* (Third Edit). AMSTERDAM • BOSTON • HEIDELBERG • LONDON NEW YORK • OXFORD • PARIS • SAN DIEGO SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO: Elsevier Ltd. Retrieved from http://www.igi-global.com/chapter/digital-evidence-computer-crime/28509

Casey, E. (2012). Cloud computing and digital forensics. *Digital Investigation*, *9*(2), 69–70. http://doi.org/10.1016/j.diin.2012.11.001

Casey, E., & Schatz, B. (2011). Digital Investigations 2.

Catryna, B. (2011). *Review of the Cybercrime Legislation Amendment Bill*. Retrieved from file:///C:/Users/salqahtany/Desktop/http---www.aphref.aph.gov.au-house-committee-jscc-cybercrime_bill-report-fullreport.pdf

Chen, G., Du, Y., Qin, P., & Du, J. (2012). Suggestions to digital forensics in Cloud computing ERA. In *2012 3rd IEEE International Conference on Network Infrastructure and Digital Content* (pp. 540–544). Beijing: Ieee. http://doi.org/10.1109/ICNIDC.2012.6418812

CHEN, H.-Y. (2014). Cloud Crime to Traditional Digital Forensic Legal and Technical Challenges and Countermeasures. In *2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)* (pp. 990–994). Ieee.

Chow, R., Golle, P., Jakobsson, M., Shi, E., Staddon, J., Masuoka, R., & Molina, J. (2009). Controlling Data in the Cloud : Outsourcing Computation without Outsourcing Control, 85–90.

Clarke, N. (2010). *Computer Forensics* (1st ed.). Cambridgeshire: IT Governance Publishing.

Cloud Industry Forum. (2012). UK Cloud adoption and trends for 2013. Retrieved July 22, 2013, from http://www.cloudindustryforum.org/white-papers/uk-cloud-adoption-and-trends-for-2013

Cloud Security Alliance. (2009). Security Guidance For Critical Areas of Focus in Cloud Computing. *Security Guidance for Critical Areas of Focus in Cloud Computing*, *2*(December), 1–76.

Cohen, M. I. (2008). PyFlag – An advanced network forensic framework. *Digital Investigation*,

5, S112–S120. http://doi.org/10.1016/j.diin.2008.05.016

ComputerWeekly.com. (2013). Top 10 cloud computing stories of 2013. Retrieved November 17, 2014, from http://www.computerweekly.com/news/2240210997/Top-10-cloud-computing-stories-of-2013

Crosbie, M. (2013). Hack the Cloud: Ethical Hacking and Cloud Forensics. In *Cybercrime and cloud forensics* (p. 17). USA: IGI Global. http://doi.org/10.4018/978-1-4666-2662-1.ch002

CSA. (2016). *Cloud Computing Top Threats in 2016 The Treacherous 12*. Retrieved from https://downloads.cloudsecurityalliance.org/assets/research/top-threats/Treacherous-12_Cloud-Computing_Top-Threats.pdf

Daly, J. (2012). Why Crime as a Service Is the Next Big Cybersecurity Threat | StateTech Magazine. Retrieved July 28, 2013, from http://www.statetechmagazine.com/article/2012/11/why-crime-service-next-big-cybersecurity-threat

Damshenas, M., Dehghantanha, A., & Mahmoud, R. (2014). A Survey on Digital Forensics Trends. *International Journal of Cyber-Security and Digital Forensics*, *3*(4), 209–234.

Damshenas, M., Dehghantanha, A., Mahmoud, R., & Shamsuddin, S. (2012). Forensics Investigation Challenges in Cloud Computing Environments. In *Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on* (pp. 190–194). Kuala Lumpur: Ieee. http://doi.org/10.1109/CyberSec.2012.6246092

Dan, D. (2009). Feature - Grid sails to the aid of shipbuilding | iSGTW. Retrieved from http://www.isgtw.org/feature/feature-grid-sails-aid-shipbuilding

Daryabar, F., Dehghantanha, A., Udzir, N. I., & Fazlida, N. (2013). A Survey About Impacts of Cloud Computing on Digital Forensics. *International Journal of Cyber-Security and Digital Forensics (IJCSDF) 2(2): 77-94 The Society of Digital Information and Wireless Communications*, *2*(2), 77–94.

Degnan, D. (2011). Accounting for the Costs of Electronic Discovery. Retrieved November 12, 2016, from http://docplayer.net/597605-Accounting-for-the-costs-of-electronic-discovery.html

Dell. (2014). Dell Reveals Insights Driving Midsize Organizations' Adoption of Security, Cloud, Mobility and Big Data. Retrieved February 15, 2014, from http://www.dell.com/learn/us/en/uscorp1/secure/2014-11-04-dell-global-technology-adoption-index

Delport, W., Olivier, M. S., & Kohn, M. (2011). Isolating a Cloud Instance for a Digital Forensic. In *ISSA*.

DFRWS. (2006). Survey of Disk Image Storage Formats. In *the 6th annual digital forensic research workshop* (pp. 1–18). New Orleans.

Donnell, J. (2016). Challenges of Data Classification. Retrieved December 6, 2016, from https://www.schellmanco.com/blog/challenges-of-data-classification

Dykstra, J., & Sherman, A. T. (2012). Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques. *Digital Investigation*, *9*, S90–S98. http://doi.org/10.1016/j.diin.2012.05.001

Dykstra, J., & Sherman, A. T. (2013). Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform. *Digital Investigation*, *10*, S87–S95.

http://doi.org/10.1016/j.diin.2013.06.010

Dykstra, J., & Sherman, A. T. A. (2011). UNDERSTANDING ISSUES IN CLOUD FORENSICS : TWO HYPOTHETICAL CASE STUDIES. In *Proceedings of the 2011 ADFSL Conference on Digital Forensics Security and Law* (pp. 1–10). Retrieved from http://www.cisa.umbc.edu/papers/dykstra-case-studies-2011.pdf

Edwards, G. (2015). INVESTIGATING IN THE CLOUD. *ACFE*.

eSecurity. (2014). Cost of Cybercrime in U.S. Retrieved November 26, 2016, from http://www.esecurityplanet.com/network-security/cost-of-cybercrime-in-u.s.-reaches-12.7-million-per-organization.html

Evans, A., Williams, A., & Graham, J. (2011). Future of Digital Forensics : A Survey of Available Training.

Fei, B. (2007). *Data visualisation in Digital Forensics*. University of Pretoria.

Florentine, S. (2016). Cloud adoption soars, but integration challenges remain. Retrieved November 10, 2016, from http://www.cio.com/article/3018156/cloud-computing/cloud-adoption-soars-but-integration-challenges-remain.html

Forenisc Focus. (2015). Tackle the Legal Issues of Obtaining Digital Evidence in the Cloud. Retrieved November 28, 2016, from http://www.forensicfocus.com/c/aid=125/webinars/2015/tackle-the-legal-issues-of-obtaining-digital-evidence-in-the-cloud/

Galante, J., Kharif, O., & Alpeyev, P. (2011). Business & Technology | PlayStation security breach shows Amazon's cloud appeal for hackers | Seattle Times Newspaper. Retrieved

July 22, 2013, from http://seattletimes.com/html/businesstechnology/2015071863_amazoncloudhackers17.html

Gartner. (2012). Gartner Says the Personal Cloud Will Replace the Personal Computer as the Center of Users' Digital Lives by 2014. Retrieved July 19, 2013, from http://www.gartner.com/newsroom/id/1947315

Gartner. (2016). Gartner Says by 2020 "Cloud Shift" Will Affect More Than $1 Trillion in IT Spending. Retrieved November 23, 2016, from http://www.gartner.com/newsroom/id/3384720

Geethakumari, G., & Belorkar, A. (2012). Regenerating Cloud Attack Scenarios using LVM2 based System Snapshots for Forensic Analysis. *International Journal of Cloud Computing and Services Science (IJ-CLOSER)*, *1*(3), 134–141.

Gonsowski, D. (2012). E-discovery costs: Pay now or pay later. Retrieved November 12, 2016, from http://www.insidecounsel.com/2012/05/23/e-discovery-costs-pay-now-or-pay-later

Grispos, G. (2012). Calm Before the Storm : The Challenges of Cloud Computing in Digital Forensics, *4*(November), 28–48.

Grispos, G., Glisson, W., & Storer, T. (2011). Calm before the Storm: The Emerging Challenges of Cloud Computing in Digital Forensics. *Dcs.gla.ac.uk*, 1–38. Retrieved from http://www.dcs.gla.ac.uk/~tws/papers/grispos11calm-rev2425.pdf

Guo, H., Jin, B., & Shang, T. (2012). Forensic Investigations in Cloud Environments. In *2012 International Conference on Computer Science and Information Processing (CSIP)* (pp. 248–251). Xi'an, Shaanxi: Ieee. http://doi.org/10.1109/CSIP.2012.6308841

Higginbotham, S. (2010). Ericsson CEO Predicts 50 Billion Internet Connected Devices by 2020. Retrieved February 22, 2014, from http://gigaom.com/2010/04/14/ericsson-sees-the-internet-of-things-by-2020/

Hooper, C., Martini, B., & Choo, K.-K. R. (2013). Cloud computing and its implications for cybercrime investigations in Australia. *Computer Law & Security Review*, *29*(2), 152–163. http://doi.org/10.1016/j.clsr.2013.01.006

Huber, M., Schrittwieser, S., Mulazzani, M., Wondracek, G., Leithner, M., & Weippl, E. (2011). Social Snapshots : Digital Forensics for Online Social Networks. In *The Annual Computer Se-curity Applications Conference*.

Hunt, R., & Zeadally, S. (2012). Network Forensics: An Analysis of Techniques, Tools, and Trends, 36–43.

IBM. (2012, April 1). IBM - Networking Concepts and Tools for IBM SmartCloud Enterprise. Retrieved May 11, 2014, from https://www-304.ibm.com/software/brandcatalog/ismlibrary/details?catalog.label=1TW10OE01

Ieong, R. S. (2006). DIGITAL FORENSIC RESEARCH FORZA : Digital Forensics Investigation Framework That Incorporate Legal Issues By Ricci Sze-Chung Ieong FORZA – Digital forensics investigation framework that incorporate legal issues. In *DFRWS*. Lafayette. http://doi.org/10.1016/j.diin.2006.06.004

InfoWorld. (2011). Top 10 benefits of server virtualization. Retrieved November 26, 2016, from http://www.infoworld.com/article/2621446/server-virtualization/server-virtualization-top-10-benefits-of-server-virtualization.html

InSecPro. (2014). Statistic and Trends. Retrieved November 18, 2014, from

http://www.insecpro.com/index.php/articles/cyber-crime-statistics

INTEL. (2013). *Virtualization and Cloud Computing*.

Internet Live State. (2016). Internet Users. Retrieved November 26, 2016, from http://www.internetlivestats.com/internet-users/

Jadeja, Y. (2012). Cloud Computing - Concepts , Architecture and Challenges (pp. 877–880).

James, J. I., Shosha, A. F., & Gladyshev, P. (2012). Digital Forensic Investigation and Cloud Computing. In *Cybercrime and cloud forensics* (pp. 1–41). http://doi.org/10.4018/978-1-4666-2662-1.ch001

Josshua, G. (2012). Protection in the cloud: Risk management and insurance for cloud computing. *Internet Law*, *15*(12).

Kent, K., Chevalier, S., Grance, T., & Dang, H. (2006). *Guide to Integrating Forensic Techniques into Incident Response*.

Kessel, P. (2014). *Security Operations Centers — helping you get ahead of cybercrime*.

Kevin, B. (2006). The Blue Pill virtualization attack: Virtual machine malware. Retrieved from http://searchservervirtualization.techtarget.com/tip/The-Blue-Pill-virtualization-attack-Virtual-machine-malware

Khajeh-hosseini, A., & Greenwood, D. (2010). Cloud Migration : A Case Study of Migrating an Enterprise IT System to IaaS Ian Sommerville. Miami, USA: IEEE.

Knox, M. (2012). Crime Scene Journal. Retrieved from http://www.crimescenejournal.com/content.php?id=0007

Ko, R. K. L., Jagadpramana, P., Mowbray, M., Pearson, S., Kirchberg, M., Liang, Q., & Lee, B. S. (2011). TrustCloud: A Framework for Accountability and Trust in Cloud Computing. In *2011 IEEE World Congress on Services* (pp. 584–588). Washington, DC: Ieee. http://doi.org/10.1109/SERVICES.2011.91

Koen, R. (2009). *The development of an open-source forensics platform*. University of Pretoria.

Krebs on Security. (2013). pavel vrublevsky — Krebs on Security. Retrieved July 23, 2013, from https://krebsonsecurity.com/tag/pavel-vrublevsky/

Kumar, M. (2011). Computer Investigations. Retrieved from http://thehackernews.com/2011/09/offline-windows-analysis-and-data.html

Kuyoro, S. O., Ibihunle, F., & Awodele, O. (2011). Cloud Computing Security Issues and Challenges. *International Journal of Computer Networks*, *3*(3), 247–255.

Law, F., & Chow, K. P. (2014). UNDERSTANDING COMPUTER FORENSICS REQUIREMENTS IN CHINA VIA THE " PANDA BURNING INCENSE " VIRUS CASE. *2014 ADFSL*, 51–58.

Lee, J., & Un, S. (2012). Digital forensics as a service: A case study of forensic indexed search. In *ICT Convergence (ICTC), 2012 International Conference* (pp. 499–503). Jeju Island.

Lewelling, M. (2013). McAfee: Cybercrime Costs Companies $100B -- Or 500,000 Jobs -- A Year. *CRN*. Retrieved from http://www.crn.com/news/security/240158985/mcafee-cybercrime-costs-companies-100b-or-500-000-jobs-a-year.htm?itc=refresh

Li, J., Chen, X., Huang, Q., & Wong, D. S. (2013). Digital provenance: Enabling secure data forensics in cloud computing. *Future Generation Computer Systems*.

http://doi.org/10.1016/j.future.2013.10.006

Lillard, T. V. (2010). *Digital forensics for network, Internet, and cloud computing a forensic evidence guide for moving targets and data*. Burlington, MA: Syngress Publishing.

Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., & Leaf, D. (2011). *NIST Cloud Computing Reference Architecture Recommendations of the National Institute of Standards and*.

Lu, R., Lin, X., Liang, X., & Shen, X. S. (2010). Secure Provenance : The Essential of Bread and Butter of Data Forensics in Cloud Computing. In *In Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security* (pp. 282–292). New York, New York, USA.

Marangos, N., Rizomiliotis, P., & Mitrou, L. (2014). Time Synchronization : Pivotal Element in Cloud Forensics. *Security and Communication Networks*.

Martini, B., & Choo, K.-K. R. (2012). An integrated conceptual digital forensic framework for cloud computing. *Digital Investigation*, *9*(2), 71–80. http://doi.org/10.1016/j.diin.2012.07.001

Marty, R. (2011). Cloud application logging for forensics. *Proceedings of the 2011 ACM Symposium on Applied Computing - SAC '11*, 178. http://doi.org/10.1145/1982185.1982226

McKemmish, R. (1999). *What is forensic computing?* Retrieved from http://aic.gov.au/documents/9/C/A/%7B9CA41AE8-EADB-4BBF-9894-64E0DF87BDF7%7Dti118.pdf

Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing, Recommendations of the National Institute of Standards and Technolog*. *National Institute of Standards and Technology*. Gaithersburg, MD. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+NIST+Definition+of+Cloud+Computing+Recommendations+of+the+National+Institute+of+Standards+and+Technology#4

Metropolitan Police. Total policing (2014).

Miller, C., Glendowne, D., Dampier, D., & Blaylock, K. (2014). Forensicloud: An Architecture for Digital Forensic Analysis in the Cloud. *Journal of Cyber Security and Mobility*, *3*(3), 231–262. http://doi.org/10.13052/jcsm2245-1439.331

Mohay, G., Anderson, A., Collie, B., De vel, O., & McKemmish, R. (2003). *Computer and intrusion forensics*. Boston. Retrieved from http://books.google.com/books?hl=en&lr=&id=z4GLgpwsYrkC&oi=fnd&pg=PR11&dq=Computer+and+Intrusion+Forensics&ots=luByAU98w_&sig=PCg7NE8lBCx3GkN95F3UuTnsx3o

Motahari-Nezhad, H. (2009). Outsourcing business to cloud computing services: Opportunities and challenges. *IEEE Internet Computing, Special Issue on Cloud Computing*. Retrieved from http://www.lrr.in.tum.de/~gerndt/home/Teaching/CloudComputing/20111006112649503.pdf

Murphy, B. (2011). e-Discovery in The Cloud Not As Simple As You Think. Retrieved September 11, 2014, from http://www.forbes.com/sites/jasonvelasco/2011/11/29/e-discovery-in-the-cloud-not-as-simple-as-you-think/

NIST. (2011). Challenging Security Requirements for US Government Cloud Computing Adoption (Draft).

Osborne, G., Turnbull, B., & Slay, J. (2010). The "Explore , Investigate and Correlate" (EIC) conceptual framework for digital forensics Information Visualisation. In *2010 International Conference on Availability, Reliability and Security* (pp. 629–634).

Oxford Dictionary. (2014). forensic: definition of forensic in Oxford dictionary (British & World English). Retrieved April 29, 2014, from http://www.oxforddictionaries.com/definition/english/forensic?q=forensics#forensic__8

Palmer, G. (2001). *A Road Map for Digital Forensic Research*. New York, New York, USA.

Paraben. (2013). *Computer Forensic Tools Comparison Chart*.

Patel, F. (2014). Comparative Study of Grid and Cloud Computing Farheen Patel Computer science. *International Journal of Sceientific Reserach*, (2277), 3–4.

Pearson, A. (2014). What's The Difference Between Hashing and Encrypting? Retrieved November 26, 2016, from http://www.securityinnovationeurope.com/blog/whats-the-difference-between-hashing-and-encrypting

Phillippi, N. (2010). The Difference Between Virtualization and Cloud Computing. Retrieved from http://www.erpsoftwareblog.com/2010/08/the-difference-between-virtualization-and-cloud-computing/

Poisel, R., Malzer, E., & Tjoa, S. (2012). Evidence and Cloud Computing : The Virtual Machine Introspection Approach. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 4*(August), 135–152.

Ponemon Institute. (2014). THE COST OF CYBER SECURITY CRIME. Retrieved November 18, 2014, from http://www8.hp.com/us/en/software-solutions/ponemon-cyber-security-report/?jumpid=ba_r329_hhoaffiliate&aid=38293&pbid=je6NUbpObpQ&aoid=35252&siteid=je6NUbpObpQ-khv19XFJmAt8emY62ihmhw

Posey, B., & Burton, A. (2015). Data Retention Policy. Retrieved December 8, 2016, from http://searchdatabackup.techtarget.com/definition/data-retention-policy

Povar, D., & Bhadran, V. K. (2010). Forensic Data carving.

Purba, N. (2016). DDoS attacks "consistent, relentless and damaging" to organizations. Retrieved December 8, 2016, from http://www.welivesecurity.com/2016/10/05/ddos-attacks-consistent-relentless-damaging-organizations/

Quick, D., & Choo, K.-K. R. (2013). Google Drive: Forensic analysis of data remnants. *Journal of Network and Computer Applications*, 1–15. http://doi.org/10.1016/j.jnca.2013.09.016

Raghavan, S. (2012). Digital forensic research: current state of the art. *CSI Transactions on ICT*, *1*(1), 91–114. http://doi.org/10.1007/s40012-012-0008-7

Red Eye Monitor. (2011). Utility Computing vs. Cloud Computing. Retrieved May 3, 2017, from https://redeyemon.wordpress.com/2010/02/07/the-difference-between-utility-and-cloud-computing/

Reilly, D., Wren, C., & Berry, T. (2011). Cloud Computing : Pros and Cons for Computer Forensic Investigations. *International Journal Multimedia and Image Processing*, *1*(1), 26–34.

Roussev, V. (2009). Hashing and Data Fingerprinting in Digital Forensics. *Ieee Computer Society*, (April), 49–55.

Ruan, K. (2013). Designing a Forensic-Enabling Cloud Ecosystem. In *Cybercrime and cloud forensics* (pp. 331–344). USA: IGI Global.

Ruan, K., & Carthy, J. (2013). Cloud Computing Reference Architecture and Its Forensic Implications: A Preliminary Analysis. *Digital Forensics and Cyber Crime*, 1–21. Retrieved from http://cloudforensicsresearch.org/publication/2012_NIST Cloud Architecture and Forensic Implications_4thICF2C_Springer.pdf

Ruan, K., Carthy, J., Kechadi, T., & Baggili, I. (2013). Cloud forensics definitions and critical criteria for cloud forensic capability: An overview of survey results. *Digital Investigation*, *10*(1), 34–43. http://doi.org/10.1016/j.diin.2013.02.004

Ruan, K., Carthy, J., Kechadi, T., & Crosbie, M. (2011). Cloud forensics : An overview. *Advances in Digital Forensics VII*, 15–26.

Sang, T. (2013). A Log Based Approach to Make Digital Forensics Easier on Cloud Computing. In *2013 Third International Conference on Intelligent System Design and Engineering Applications* (pp. 91–94). Ieee. http://doi.org/10.1109/ISDEA.2012.29

Shah, J. J., & Malik, L. G. (2013). Cloud Forensics: Issues and Challenges. *2013 6th International Conference on Emerging Trends in Engineering and Technology*, 138–139. http://doi.org/10.1109/ICETET.2013.44

Sibiya, G., Venter, H. S., & Fogwill, T. (2012). Digital Forensic Framework for a Cloud Environment. In *IST_Africa 2012 Conference proceedings* (pp. 1–8).

Singh, A. (2004). An Introduction to Virtualization. Retrieved from http://www.kernelthread.com/publications/virtualization/

Sourya, B. (2011). Cloud Computing vs Utility Computing vs Grid Computing: Sorting The Differences | CloudTweaks. Retrieved from http://www.cloudtweaks.com/2011/02/cloud-computing-vs-utility-computing-vs-grid-computing-sorting-the-differences/

Srinivasa, G., Shaik, N., UshaRani, U., & Krishna, G. V. (2011). A Study on the Readiness of Cloud Computing for Captious Computations. *Wcsit.org*, *1*(6), 247–252. Retrieved from http://www.wcsit.org/pub/2011/July/A Study on the Readiness of Cloud Computing for Captious Computations.pdf

Steinberg, J. (2014). Nude Photos Of Jennifer Lawrence And Kate Upton Leak: Five Important Lessons For All of Us. Retrieved September 11, 2014, from http://www.forbes.com/sites/josephsteinberg/2014/08/31/nude-photos-of-jessica-lawrence-and-kate-upton-leak-five-important-lessons-for-all-of-us/

Taylor, M., Haggerty, J., Gresty, D., & Hegarty, R. (2010). Digital evidence in cloud computing systems. *Computer Law & Security Review*, *26*(3), 304–308. http://doi.org/10.1016/j.clsr.2010.03.002

Taylor, M., Haggerty, J., Gresty, D., & Lamb, D. (2011). Forensic investigation of cloud computing systems. *Network Security*, *2011*(3), 4–10. http://doi.org/10.1016/S1353-4858(11)70024-1

Thethi, N., & Keane, A. (2014). Digital forensics investigations in the Cloud. *2014 IEEE International Advance Computing Conference (IACC)*, 1475–1480. http://doi.org/10.1109/IAdCC.2014.6779543

thrivenetworks. (2016). What is Regulatory Compliance and Why is It Important?

Trenwith, P. M., & Venter, H. (2013). Digital Forensic Readiness in the Cloud. In *Information Security for South Africa, 2013* (pp. 1–5).

Valjarevic, A., & Venter, H. (2013). Implementation Guidelines for a Harmonised Digital Forensic Investigation Readiness Process Model. *Information Security for South Africa, 2013*, 1–9.

Vaquero, L. M., Rodero-merino, L., Caceres, J., & Lindner, M. (2009). A Break in the Clouds : Towards a Cloud Definition. *ACM SIGCOMM Computer Communication Review*, *39*(1), 50–55.

VMware. (2015). VMware Workstation Pro. Retrieved from http://www.vmware.com/products/workstation

Wang, S.-J. (2007). Measures of retaining digital evidence to prosecute computer-based cyber-crimes. *Computer Standards & Interfaces*, *29*(2), 216–223. http://doi.org/10.1016/j.csi.2006.03.008

Weise, E. (2015). Massive breach at health care company Anthem Inc. Retrieved November 22, 2016, from http://www.usatoday.com/story/tech/2015/02/04/health-care-anthem-hacked/22900925/

Wen, Y., Man, X., Le, K., & Shi, W. (2013). Forensics-as-a-Service (FaaS): Computer Forensic Workflow Management and Processing Using Cloud. In *Cloud Computing 2013* (pp. 208–214).

Wilson, P. (2011). Positive perspectives on cloud security. *Information Security Technical*

Report, *16*(3–4), 97–101. http://doi.org/10.1016/j.istr.2011.08.002

Wolthusen, S. D. (2009). Overcast: Forensic Discovery in Cloud Environments. *2009 Fifth International Conference on IT Security Incident Management and IT Forensics*, 3–9. http://doi.org/10.1109/IMF.2009.21

Yadav, S. (2011). Analysis of Digital Forensic and Investigation 1. *VSRD International Journal of Computer Science & Information Technology*, *1*(3), 171–178.

Yan, C. (2011). Cybercrime forensic system in cloud computing. In *Proceedings of 2011 International Conference on Image Analysis and Signal Processing, IASP 2011* (pp. 612–613). http://doi.org/10.1109/IASP.2011.6109117

Yen, P., Yang, C., & Ahn, T. (2009). Design and Implementation of a Live-analysis Digital Forensic System. *International Conference on Convergence and Hybrid Information Technology 2009 Design*, 1–5.

Yu, W., & Wang, C. (2011). TOWARD THE TREND OF CLOUD COMPUTING, 238–242.

Zaferullah, Z., Anwar, F., & Anwar, Z. (2011). Digital Forensics for Eucalyptus. In *2011 Frontiers of Information Technology* (pp. 110–116). Islamabad: Ieee. http://doi.org/10.1109/FIT.2011.28

Zargari, S., & Benford, D. (2012). Cloud Forensics: Concepts, Issues, and Challenges. In *2012 Third International Conference on Emerging Intelligent Data and Web Technologies* (pp. 236–243). Bucharest: Ieee. http://doi.org/10.1109/EIDWT.2012.44

ZatykoDr, K., & Bay, J. (2011). The Digital Forensics Cyber Exchange Principle. *Forensic Magazine*. Retrieved from http://www.forensicmag.com/articles/2011/12/digital-

forensics-cyber-exchange-principle

Zawoad, S., Dutta, A., & Hasan, R. (2013). SecLaaS: secure logging-as-a-service for cloud forensics. ... , *Computer and Communications Security*. Retrieved from http://dl.acm.org/citation.cfm?id=2484342

Zawoad, S., & Hasan, R. (2012). I Have the Proof : Providing Proofs of Past Data Possession in Cloud Forensics.

Zawoad, S., & Hasan, R. (2013a). Cloud Forensics: A Meta-Study of Challenges, Approaches, and Open Problems. *arXiv Preprint arXiv:1302.6312*, 1–15. Retrieved from http://arxiv.org/abs/1302.6312

Zawoad, S., & Hasan, R. (2013b). Digital Forensics in the Cloud. *CrossTalk*, (October), 17–20.

Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. *Journal of Internet Services and Applications*, *1*(1), 7–18. http://doi.org/10.1007/s13174-010-0007-6

Zimmerman, S., & Glavach, D. (2011). Cyber Forensics in the Cloud. *IANewsletter*, *14*(1), 1–36.

Appendix A – Ethical Approval

RESEARCH WITH PLYMOUTH UNIVERSITY

19 July 2016

CONFIDENTIAL
Saad Alqahtany
School of Computing, Electronics and Mathematics

Dear Saad

Ethical Approval Application

Thank you for submitting the ethical approval form and details concerning your project:

A Forensically Enabled Cloud Computing Architecture for IaaS (Infrastructure as a Service) model

I am pleased to inform you that this has been approved.

Kind regards

Paula Simson
Secretary to Faculty Research Ethics Committee

Cc. Prof Nathan Clarke
Prof Steven Furnell
Prof Christoph Reich

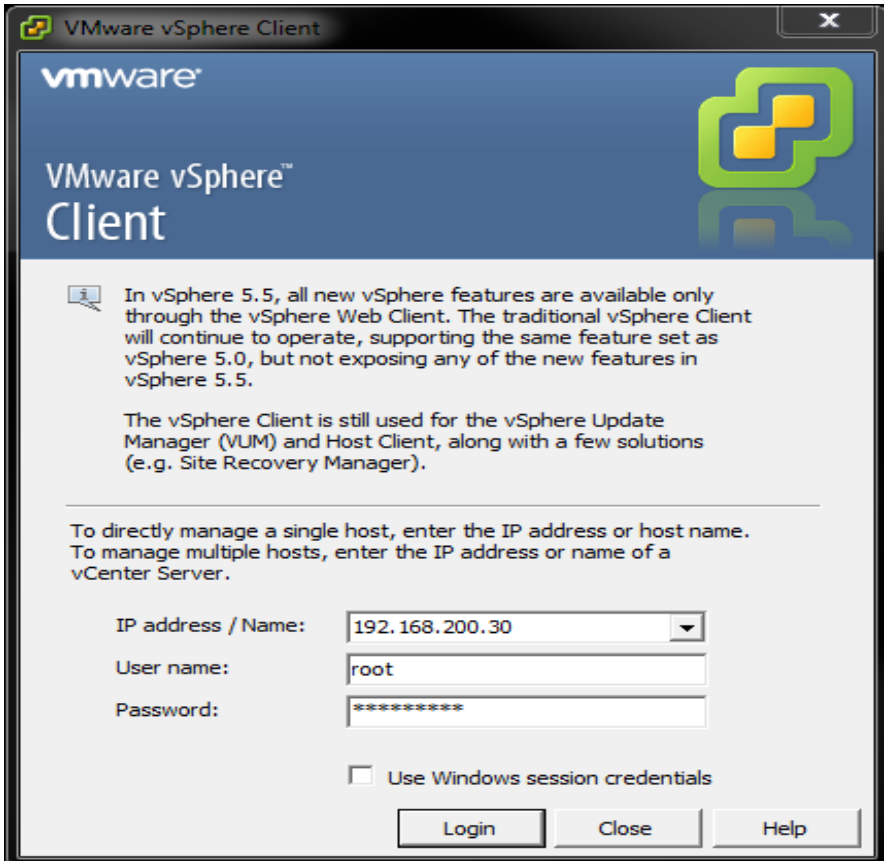
Faculty of Science and Engineering
Plymouth University
Drake Circus
PL4 8AA

T +44 (0) 1752 584 584
F +44 (0) 1752 584 540
W www.plymouth.ac.uk

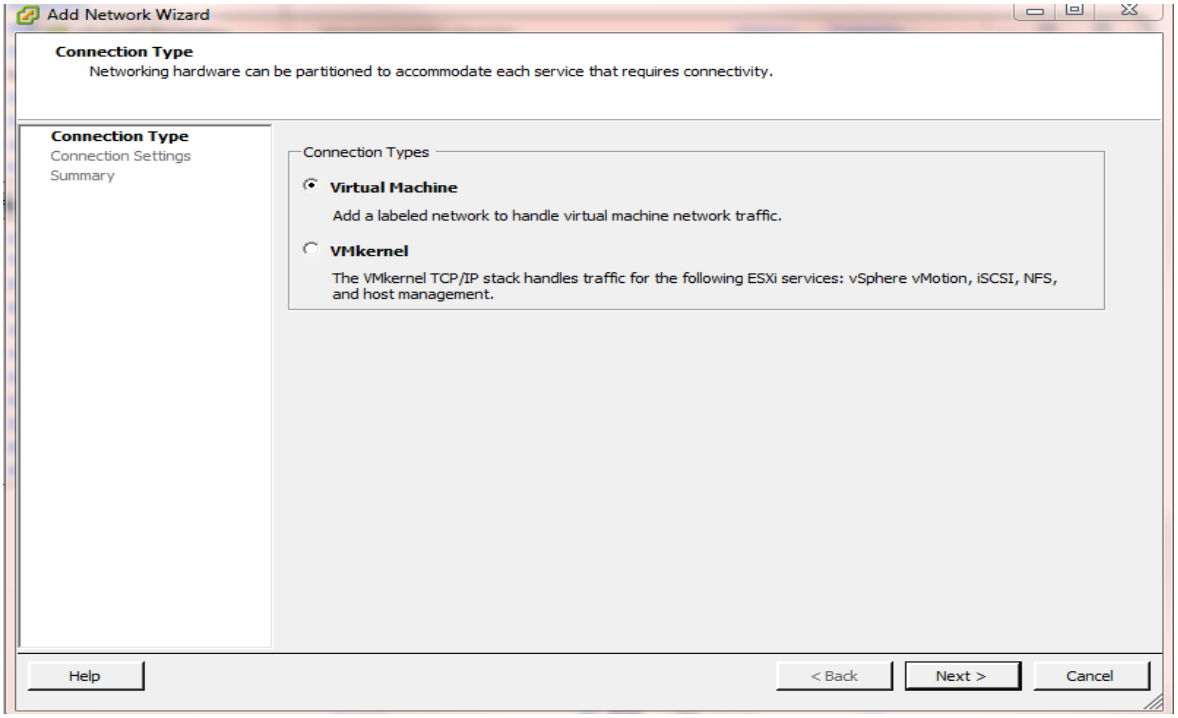
Mrs Christine Mushens BA
Faculty Business Manager

Installation Guide Documentation

1- Access the physical host via vSphere client using its IP, username and Password.



2- **Add a Virtual Machine network to handle virtual machine network traffic** for example (VMNetwork1) as following: Configuration- network- vSwitch0 properties-add-virtual Machine.



3- Add another Virtual machine network as above step (VMNetwork2)

4- Now deploy OVF vCenter appliance, Vshield Manager and Vcloud Director (File-deploy OVF template).

5- Give a static IP address to each virtual machine appliance.

6- Suggested specification of each VM

	Physical ESXi Host	Virtual (ESXi Cluster)	vCenter	vShiledMgr	VCD
RAM	65G	50 G	8 G	6 G	4 G
HD	10 T (Cloud)	4 T	125 G	60 GB	30 G
CPUs	12*2	2*6 Per Socket	2	1	1

7- Access the vCenter via vSphere client using its IP, username and Password

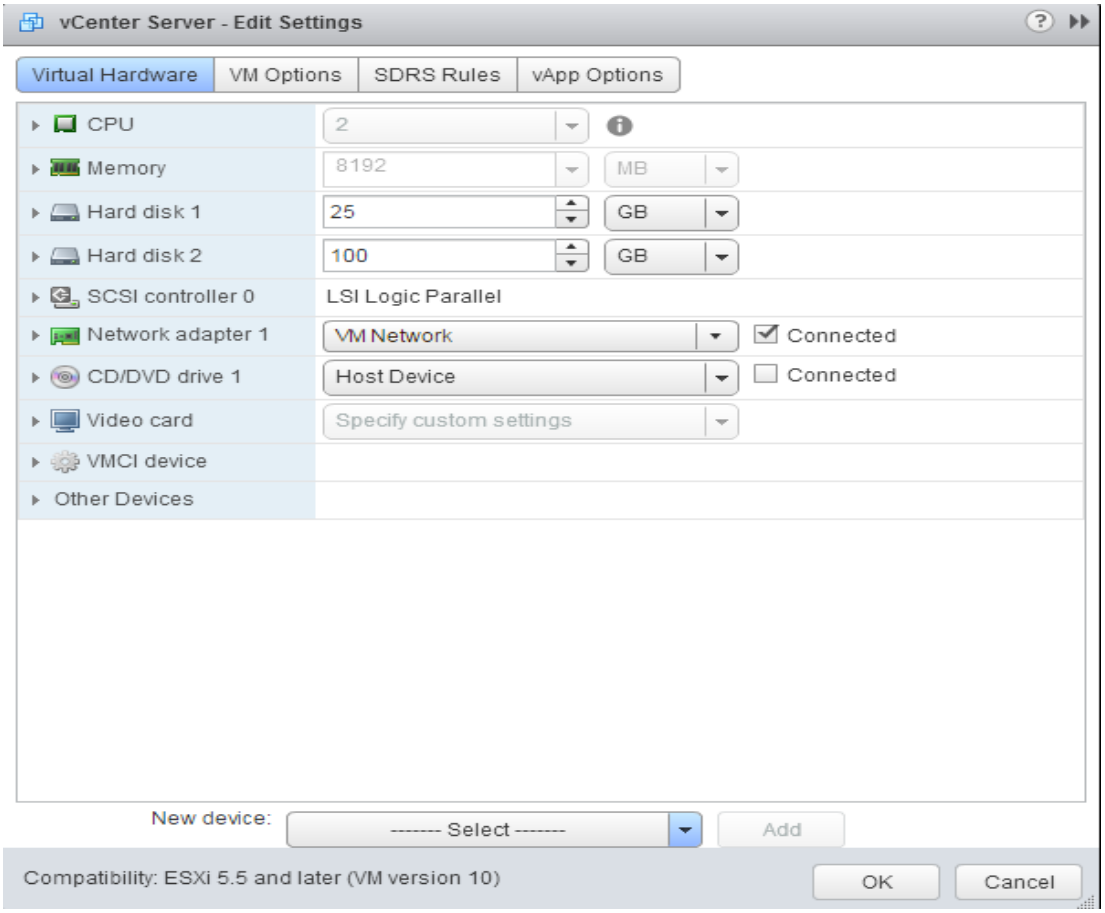
8- Create data centre.

9- Create main cluster.

10- Add the physical host to this main cluster.

11- Use vSphere web client using the port number 9443 for example https://192.168.200.56:9443

12- Edit vCenter Virtual machine:

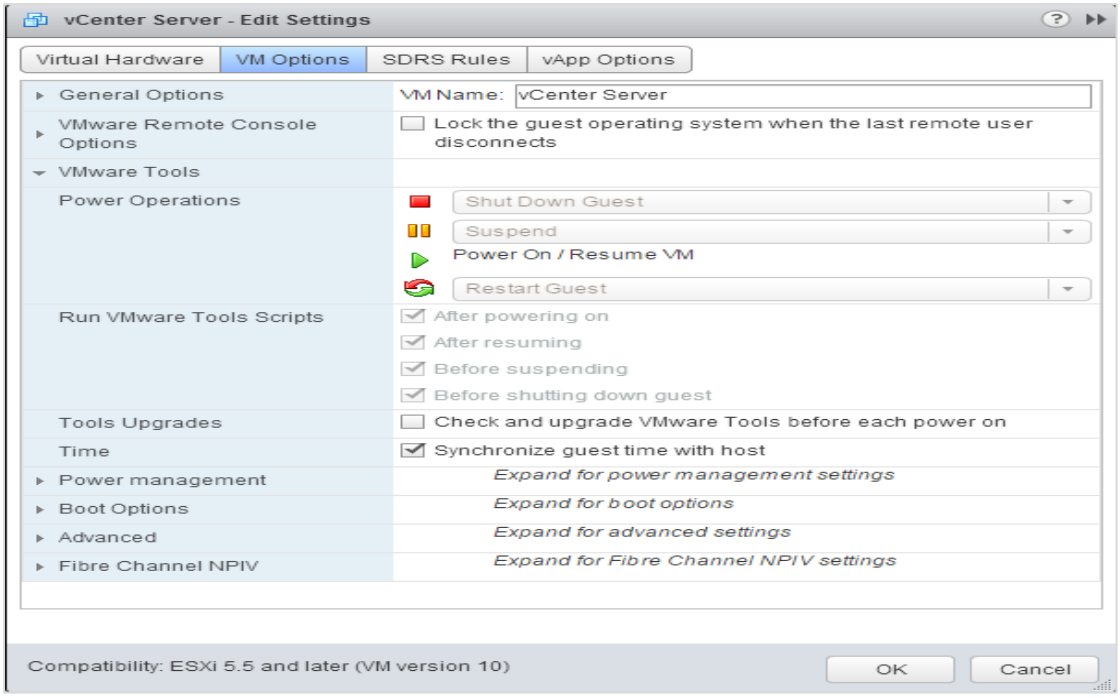


13- Add new network- select VMNetwork1 created in step 2.

14- Repeat step 12 for vShiled manger and Vcloud director machines. (Vcloud Director has to have 2 VMNetworks (1&2).

15- Please check (Upgrade VMware tools and Synchronise guest time with host) check boxes For vecenter, Vshield Mag and Vcloud director.

From VM Option- VMware Tools.



16- Via vShpere client, add new virtual machine called ESXi Host with 6 GB memory, 4 CPUs and 1 TB HD.

17- Once it is created. Edit this virtual host,CD/DVD drive one, select ESXi ISO image, check connect at Power on. Wait till ESXi installed successfully in this machine.

18- Press F2 to access it. Give it static IP address, default gateway and Net mask.

19- Press ESC to save changes

20- Create new cluster.

21- Add this virtual host to this cluster.

22- We need to add a V Sphere Distributed Switch to this machine. Otherwise, VCD cannot talk to it.

23- Access Vshild manager via its web for example:

https://192.168.200.37

User name: Admin

Password: default.

If The Web interfaces to vShield Manager does not get updated, o resolve this issue, disable and then re-enable Web services.

To disable and re-enable Web services:

Log in to the vShield Manager virtual machine using this credential:

Username: admin

Default password: default

To enable root privilege, run this command:

Enable

Enter the admin password again.

Note: Use this privileged command only under the direction of VMware Support.

Run this command to configure the terminal:

Config t

Run this command to disable the Web services:

No web-manager

Wait for a second or two and then enable the Web service using this command:

Web-manager

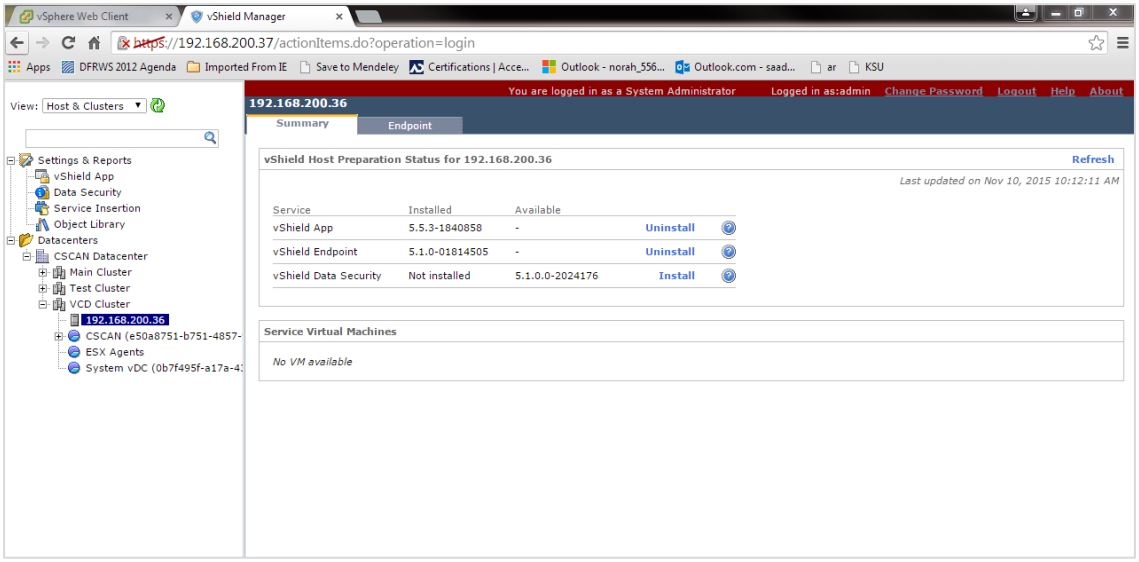
Run this command twice to exit the system:

Exit

Reload the client to see the change

24- Edit lookup service, vCenter Server, DNS servers & NTP Server.

25- Go datacentres, select the virtual host. Install vShield App and vShiled Endpoint.



26- Access VCD (vCloud Director) for example:
https://192.168.200.38

27- Go through Guided Tasks (1 to 6 steps). We do not use another network pool or catalogue. Thus please ignore step 4 & 6.

28- Once you completed these tasks, you should be able to access the could use your organisation ID and import your vApps from the main cluster.

Appendix C - Cloud FAAS software code

Based Image

```
@echo on

"format" e: /y

timeout /t 30

"dcfldd.exe" if=//./e: hash=md5 hashwindow=65536 bs=65536 md5log=hash2.txt hashconv=after > nul 2>&1

"dcfldd.exe" if=//./e: hash=md5 hashwindow=65536 bs=65536 md5log=hash1.txt hashconv=after > nul 2>&1

"dcfldd.exe" if=//./e: hash=md5 hashwindow=1024M bs=65536 hashlog=diskhash2.txt hashconv=after of=c:\temp\base.img> nul 2>&1

"dcfldd.exe" if=//./e: hash=md5 hashwindow=1024M bs=65536 hashlog=diskhash.txt hashconv=after of=c:\temp\base.img> nul 2>&1

move c:\temp\base.img c:\base
```

Monitor Changes

```
@echo on

check.py

timeout 60

backup.bat
```

Check

```
import os


import time


import linecache

import subprocess

os.system('dcfldd.exe skip=5460 count=128 if=//./e: hash=md5 hashwindow=1024M bs=65536 hashlog=diskhash2.txt hashconv=after > nul 2>&1')

linecache.clearcache()
```

```
hosts0 = linecache.getline('diskhash.txt', 1)

hosts1 = linecache.getline('diskhash2.txt', 1)

print hosts1

print hosts0

if hosts0 == hosts1:

print 0

quit()

else:

os.system('readonly.bat')

quit()
```

Read only

```
@echo on

diskpart /s script.txt

match1.py

diskpart /s script2.txt

dcfldd.exe if=//./e: skip=5460 count=128 hash=md5 hashwindow=1024M bs=65536 hashlog=diskhash2.txt hashconv=after > nul 2>&1

copy diskhash2.txt diskhash.txt /y
```

Match1

```
import os

import linecache

import subprocess
```

```
import time

os.system('dcfldd.exe if=//./e: hash=md5 hashwindow=1024M bs=65536 hashlog=diskhash2.txt hashconv=after > nul 2>&1')

hosts0 = linecache.getline('diskhash.txt', 2)

hosts1 = linecache.getline('diskhash2.txt', 2)

if hosts0 != hosts1:

print hosts1

os.system('dcfldd if=//./e: hash=md5 hashwindow=65536 bs=65536 md5log=hash2.txt hashconv=after > nul 2>&1')

hosts0 = open("hash1.txt","r")

hosts1 = open("hash2.txt","r")

lines1 = hosts0.readlines()

for i,lines2 in enumerate(hosts1):

if lines2 != lines1[i] and "Total (md5):" not in lines2:

print str(i)

subprocess.check_output("\"dcfldd.exe\" if=\\\.\\e: hash=md5 md5log=c:\\temp\\"+str(i)+".txt of=c:\\temp\\"+str(i)+".img skip="+str(i)+" count=1 bs=65536"+"\n", shell=True)


if=%cd%\\"+str(i)+".img seek="+str(i)+" count=1 bs=65536"+"\n")

with open("hash2.txt") as f:

with open("hash1.txt", "w") as f1:

for line in f:

f1.write(line)

with open("diskhash2.txt") as f:

with open("diskhash.txt", "w") as f1:

for line in f:

f1.write(line)

os.system('maketime.bat')
```

Make time

```
@echo on

CD  C:\temp

SET HOUR=%time:~0,2%

SET dtStamp24=%date:~-4%%date:~4,2%%date:~7,2%_%time:~0,2%%time:~3,2%

SET Datefolder=%dtStamp24%

mkdir C:\temp\%Datefolder%

move C:\temp\*.* C:\temp\%Datefolder%

dir C:\temp\%Datefolder%

copy h:\scan\diskhash2.txt C:\temp\%Datefolder%

copy h:\scan\hash2.txt C:\temp\%Datefolder%

copy h:\scan\restore.bat C:\temp\%Datefolder%

copy h:\scan\dd.exe C:\temp\%Datefolder%
```

Reconstruction Engine

```
if not exist c:\temp\%username%\base.img copy c:\base\%username%\base.img c:\temp\%username%
/y

for %%f in (*.img) do (

        "dd.exe"  of=c:\temp\%username%\base.img  if=%cd%\%%~nf.img  seek=%%~nf  count=1
bs=65536

   )

pause

cmd
```